

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
4 July 2002 (04.07.2002)

PCT

(10) International Publication Number  
WO 02/052015 A2(51) International Patent Classification<sup>7</sup>: C12N 15/51,  
15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85(74) Agent: BERNIER, Louise, G.; 2100 Cunard Street,  
Laval, Québec H7S 2G5 (CA).

(21) International Application Number: PCT/CA01/01843

(22) International Filing Date:  
20 December 2001 (20.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/257,857 22 December 2000 (22.12.2000) US(71) Applicant (for all designated States except US):  
BOEHRINGER INGELHEIM (CANADA) LTD.  
[CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5  
(CA).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,  
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,  
ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

(72) Inventors; and





(75) Inventors/Applicants (for US only): KUKOLJ, George  
[CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5  
(CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street,  
Laval, Québec H7S 2G5 (CA).

Published:

— without international search report and to be republished  
upon receipt of that report

[Continued on next page]

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

CLONE APGK-12		AMINO ACID SUBSTITUTIONS								
5' HCV IRES		NeoR	EMCV IRES	NS2	NS3	HCV NS2→5B 4A	NS4B	NS5A	NS5B	3'HCV UTR
 77 cfu/μg	G (nt1) SEQ ID NO 1									
 66 cfu/μg	A (nt1) SEQ ID NO 24			-	-	-	-	-	-	
 1100000cfu/μg	R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1691)R	-	T(1993)A G(2042)C L(2155)P P(2166)L		
 2000000cfu/μg	G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1691)R	-	T(1993)A G(2042)C L(2155)P P(2166)L		

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G→A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

### FIELD OF THE INVENTION

The present invention relates generally to a HCV RNA molecule that self-replicates  
5 in appropriate cell lines, particularly to a self-replicating HCV RNA construct having  
an enhanced efficiency of establishing cell culture replication.

### BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion and  
10 community-acquired non-A non-B hepatitis worldwide. It is estimated that over 200  
million people worldwide are infected by the virus. A high percentage of carriers  
become chronically infected and many progress to chronic liver disease, so called  
chronic hepatitis C. This group is in turn at high risk for serious liver disease such as  
liver cirrhosis, hepatocellular carcinoma and terminal liver disease leading to death.  
15 The mechanism by which HCV establishes viral persistence and causes a high rate  
of chronic liver disease has not been thoroughly elucidated. It is not known how  
HCV interacts with and evades the host immune system. In addition, the roles of  
cellular and humoral immune responses in protection against HCV infection and  
disease have yet to be established.

20 Various clinical studies have been conducted with the goal of identifying  
pharmaceutical compounds capable of effectively treating HCV infection in patients  
afflicted with chronic hepatitis C. These studies have involved the use of interferon-  
alpha, alone and in combination with other antiviral agents such as ribavirin. Such  
25 studies have shown that a substantial number of the participants do not respond to  
these therapies, and of those that do respond favorably, a large proportion were  
found to relapse after termination of treatment. To date there are no broadly  
effective antiviral compounds for treatment of HCV infection.

30 HCV is an enveloped positive strand RNA virus in the *Flaviviridae* family. The single  
strand HCV RNA genome is of positive polarity and comprises one open reading  
frame (ORF) of approximately 9600 nucleotides in length, which encodes a linear  
polyprotein of approx. 3010 amino acids. In infected cells, this polyprotein is cleaved  
at multiple sites by cellular and viral proteases to produce structural and non-  
35 structural (NS) proteins. The structural proteins (C, E1, E2 and E2-p7) comprise

## 2

polypeptides that constitute the virus particle (Hijikata *et al.*, 1991; Grakoui *et al.*, 1993(a) ). The non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B) encode for enzymes or accessory factors that catalyze and regulate the replication of the HCV RNA genome. Processing of the structural proteins is catalyzed by host cell proteases (Hijikata *et al.*, 1991). The generation of the mature non-structural proteins is catalyzed by two virally encoded proteases. The first is the NS2/3 zinc-dependent metalloprotease which auto-catalyses the release of the NS3 protein from the polyprotein. The released NS3 contains a N-terminal serine protease domain (Grakoui *et al.*, 1993(b); Hijikata *et al.*, 1993) and catalyzes the remaining cleavages from the polyprotein. The released NS4A protein has at least two roles. First, forming a stable complex with NS3 protein and assisting in the membrane localization of the NS3/NS4A complex (Kim *et al.*, Arch Virol. 1999, 144: 329-343) and second, acting as a cofactor for NS3 protease activity. This membrane-associated complex, in turn catalyzes the cleavage of the remaining sites on the polyprotein, thus effecting the release of NS4B, NS5A and NS5B (Bartenschlager *et al.*, 1993; Grakoui *et al.*, 1993(a); Hijikata *et al.*, 1993; Love *et al.*, 1996; reviewed in Kwong *et al.*, 1998). The C-terminal segment of the NS3 protein also harbors nucleoside triphosphatase and RNA helicase activity (Kim *et al.*, 1995). The function of the protein NS4B is unknown. NS5A, a highly phosphorylated protein, seems to be responsible for the Interferon resistance of various HCV genotypes (Gale Jr. *et al.*, 1997 Virology 230, 217; Reed *et al.*, 1997. NS5B is an RNA-dependent RNA polymerase (RdRp) that is involved in the replication of HCV.

The open reading frame of the HCV RNA genome is flanked on its 5' end by a non-translated region (NTR) of approx. 340 nucleotides that functions as the internal ribosome entry site (IRES), and on its 3' end by a NTR of approximately 230 nucleotides. Both the 5' and 3' NTRs are important for RNA genome replication. The genomic sequence variance is not evenly distributed over the genome and the 5'NTR and parts of the 3'NTR are the most highly conserved portions. The authentic, highly conserved 3'NTR is the object of US patent 5,874,565 granted to Rice *et al.*

The cloned and characterized partial and complete sequences of the HCV genome have also been analyzed with regard to appropriate targets for a prospective antiviral therapy. Four viral enzyme activities provide possible targets such as (1) the NS2/3 protease; (2) the NS3/4A protease complex, (3) the NS3 Helicase and (4) the NS5B

RNA-dependent RNA polymerase. The NS3/4A protease complex and the NS3 helicase have already been crystallized and their three-dimensional structure determined (Kim *et al.*, 1996; Yem *et al.*, 1998; Love *et al.*, 1996; Kim *et al.*, 1998; Yao *et al.*, 1997; Cho *et al.*, 1998). The NS5B RNA dependent RNA polymerase has  
5 also been crystallized to reveal a structure reminiscent of other nucleic acid polymerases (Bressanelli *et al.* 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039; Ago *et al.* 1999, Structure 7: 1417-1426; Lesburg *et al.* 1999, Nat. Struct. Biol. 6: 937-943).

10 Even though important targets for the development of a therapy for chronic HCV infection have been defined with these enzymes and even though a worldwide intensive search for suitable inhibitors is ongoing with the aid of rational drug design and HTS, the development of therapy has one major deficiency, namely the lack of  
15 cell culture systems or simple animal models, which allow direct and reliable propagation of HCV viruses. The lack of an efficient cell culture system is still the main reason to date that an understanding of HCV replication remains elusive.

Although flavi- and pestivirus self-replicating RNAs have been described and used for the replication in different cell lines with a relatively high yield, similar experiments  
20 with HCV have not been successful to date (Khromykh *et al.*, 1997; Behrens *et al.*, 1998; Moser *et al.*, 1998). It is known from different publications that cell lines or primary cell cultures can be infected with high-titer patient serum containing HCV (Lanford *et al.* 1994; Shimizu *et al.* 1993; Mizutani *et al.* 1996; Ikeda *et al.* 1998; Fournier *et al.* 1998; Ito *et al.* 1996). However, these virus-infected cell lines or cell  
25 cultures do not allow the direct detection of HCV-RNA or HCV antigens.

It is also known from the publications of Yoo *et al.* 1995; and of Dash *et al.*, 1997; that hepatoma cell lines can be transfected with synthetic HCV-RNA obtained through *in vitro* transcription of the cloned HCV genome. In both publications the  
30 authors started from the basic idea that the viral HCV genome is a plus-strand RNA functioning directly as mRNA after being transfected into the cell, permitting the synthesis of viral proteins in the course of the translation process, and so new HCV particles could form HCV viruses and their RNA detected through RT-PCR. However the published results of the RT-PCR experiments indicate that the HCV  
35 replication in the described HCV transfected hepatoma cells is not particularly

efficient and not sufficient to measure the quality of replication, let alone measure the modulations in replication after exposure to potential antiviral drugs. Furthermore it is now known that the highly conserved 3' NTR is essential for the virus replication (Yanagi *et al.*, 1999). This knowledge strictly contradicts the statements of Yoo *et al.* 5 (*supra*) and Dash *et al.* (*supra*), who used for their experiments only HCV genomes with shorter 3' NTRs and not the authentic 3' end of the HCV genome.

In WO 98/39031, Rice *et al.* disclosed authentic HCV genome RNA sequences, in particular containing: a) the highly conserved 5'-terminal sequence "GCCAGCC"; b) 10 the HCV polyprotein coding region; and c) 3'-NTR authentic sequences.

In WO 99/04008, Purcell *et al.* disclosed an HCV infectious clone that also contained only the highly conserved 5'-terminal sequence "GCCAGC".

15 Recently Lohman *et al.* 1999 (Science 285: 110-113) and Bartenschlager *et al.* (in CA 2,303,526, laid-open on October 3, 2000) disclosed a HCV cell culture system where the viral RNA (I377/NS2-3') self-replicates in the transfected cells with such efficiency that the quality of replication can be measured with accuracy and reproducibility. The Lohman and Bartenschlager disclosures were the first 20 demonstration of HCV RNA replication in cell culture that was substantiated through direct measurement by Northern blots. This replicon system and sequences disclosed therein highlight once again the conserved 5' sequence "GCCAGC". A similar observation highlighting the conservation of the 5'NTR was made by Blight *et al.* 2000 (Science 290: 1972-1974) and WO 01/89364 published on Nov. 29, 2001.

25 In addition to the conservation of the 5' and 3' untranslated regions in cell culture replicating RNAs, three other publications by Lohman *et al.* 2001, Krieger *et al.* 2001 and Guo *et al.* 2001 have recently disclosed distinct adaptive mutants within the HCV non-structural protein coding region. Specific nucleotide changes that alter the 30 amino acids of the HCV non-structural proteins are shown to enhance the efficiency of establishing stable replicating HCV subgenomic replicons in culture cells.

Applicant has now found that, contrary to all previous reports, the highly conserved 5'-NTR can be mutated by adaptation to give rise to a HCV RNA sequence that, in 35 conjunction with mutations in the HCV non-structural region, provides for a greater

efficiency of transduction and/or replication.

Applicant has also identified novel adaptive mutations within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture.

One advantage of the present invention is to provide an alternative to these existing systems comprising a HCV RNA molecule that self-replicates. Moreover, the present invention demonstrates that the initiating nucleotide of the plus-strand genome can be either an A as an alternative to the G already disclosed.

A further advantage of the present invention is to provide a unique HCV RNA molecule that transduces and/or replicates with higher efficiency. The Applicant demonstrates the utility of this specific RNA molecule in a cell line and its use in evaluating a specific inhibitor of HCV replication.

#### SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides a 5'-non translated region of the hepatitis C virus wherein its highly conserved guanine at position 1 is substituted for adenine.

Particularly, the present invention provides a hepatitis C virus polynucleotide comprising adenine at position 1 as numbered according to the I377/NS2-3' construct (Lohmann et al. 1999, Accession # AJ242651).

Particularly, the invention provides a HCV self-replicating polynucleotide comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO. 8).

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.



## 6

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

- 5 More particularly, the invention provides a HCV self-replicating polynucleotide encoding a polyprotein comprising a G2042C or a G2042R mutation.

Most particularly, the invention provides for HCV self-replicating polynucleotide comprising a nucleotide substitution G→A at position 1, and said polynucleotide  
10 encodes a polyprotein further comprising a G2042C or a G2042R mutation.

Particularly, the polynucleotide of the present invention can be in the form of RNA or DNA that can be transcribed to RNA.

- 15 In a third embodiment, the invention also provides for an expression vector comprising a DNA form of the above polynucleotide, operably linked with a promoter.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or the vector as described above.

20

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- incubating the host cell as described above in the absence or presence of a potential hepatitis C virus inhibitor;
- 25 - isolating the total cellular RNA from the cells;
- analyzing the RNA so as to measure the amount of HCV RNA replicated;
- comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.

- 30 In a sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- a) treating the above described host cell with the compound;
- b) evaluating the treated host cell for reduced replication, wherein reduced replication indicates the ability of the compound to inhibit replication.

35

**DETAILED DESCRIPTION OF THE DRAWINGS**

- Figure 1** is a schematic view of the bi-cistronic replicon RNA. The sequence deviations between the I377/NS2-3' replicon from Lohman *et al.*, 1999 and the APGK12 replicon are indicated below the replicon. In place of a G nucleotide at the +1 position in the I377/NS2-3' replicon, the APGK12 contains an additional G resulting in GG at the 5' terminus (the first G being counted as position -1). In the linker region between the neo gene and the EMCV IRES sequence two areas deviate from I377/NS2-3': 14 nucleotides (CGCGCCCAGATGTT) which are not present in I377/NS2/3 are inserted at position 1184 in APGK12; 11 nucleotides (1231-1241) present in I377/NS2-3' are deleted to generate APGK-12. In the NS5B coding region, a T at position 8032 was mutated to C to eliminate a NcoI restriction site.
- Figure 2** shows Northern blots of RNA-transfected Huh-7 cell lines. 12 µg of total cellular RNA or control RNA was separated on 0.5% agarose-formaldehyde gels and transferred to Hybond N+ paper, fixed and (Figure 2A) radioactively probed with HCV specific minus-strand RNA that detects the presence of plus-strand replicon RNA. Lanes 1 and 2: positive controls that contain 10<sup>9</sup> copies of *in vitro* transcribed APGK12 RNA. Lane 3: negative control of total cellular RNA from untransfected Huh-7 cells. Lanes 4 and 5: cellular RNA from B1 and B3 cell lines that have integrated DNA copies of the neomycin phosphotransferase gene. Lane 6: total cellular RNA from a Huh-7 cell line, designated S22.3, that harbors high copy number HCV sub-genomic replicon RNA as highlighted by the arrow. Other cell lines have no detectable replicon RNA. Figure 2B is identical to Figure 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA. Lanes 1 and 2 are positive control lanes that contain 10<sup>9</sup> copies of full length HCV minus strand RNA. Lane 6, which contains 12 µg of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8 – 9 kilobases. M represent the migration of non-radioactive molecular size markers on the agarose gel. 28s represents the migration of 28s ribosomal RNA and accounts for the detection of this species in a samples of total cellular RNA.
- Figure 3** shows indirect immunofluorescence of a HCV non-structural protein in the

S22.3 cell line. Indirect immunofluorescence was performed on cells that were cultured and fixed, permeabilized and exposed to a rabbit polyclonal antibody specific for a segment of the HCV NS4A protein. Secondary goat anti-rabbit antibody conjugated with red-fluor Alexa 594 (Molecular Probes) was used for detection. Top panels shows the results of immunofluorescence (40X objective) and the specific staining of the S22.3 cells. The bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have background level of staining.

**Figure 4** shows Western-blots following SDS-PAGE separation of total proteins extracted from three cell lines: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary \ goat anti-rabbit antibody. Panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than control B1 cells and that the naïve Huh-7 cell line does not produce the NPT protein. Panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. M represents molecular weight (in kilodaltons) of pre-stained polypeptide markers.

**Figure 5A and 5B** identify the nucleotide and amino acid sequences respectively that differ from the APGK12 sequence in the different HCV bi-cistronic replicons. The S22.3 adapted replicon is a first generation replicon selected following the transfection of RNA transcribed from the APGK12 template. R3, R7, R16 are second generation replicons that were selected following the transfection of RNA isolated from the S22.3 first generation replicon cell line. Figure 5A: Nucleotide mutations that were characterized in each of the adapted replicons are indicated adjacent to the respective segment of the replicon (IRES, NS3, NS4A, NS5A, and NS5B). Figure 5B: Amino acid numbers are numbered according to the full length HCV poly-protein with the first amino acid in the second cistron corresponding to amino acid 810 in NS2 of I377/NS2-3' construct.

**Figure 6** depicts the colony formation efficiency of four *in vitro* transcribed HCV sub-genomic bi-cistronic replicon RNAs. The APGK12 serves as the reference sequence; highlighted are the initiating nucleotides of the HCV IRES in each of the constructs and the amino acid differences (from the APGK12 reference sequence) in the HCV non-structural region for the two R3-rep. Note that the *in vitro* transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/μg in panels A and B) following selection with 0.25 mg/ml G418. RNA isolated from the second generation R3 cell line was reverse transcribed into DNA and cloned into the pAPGK12 vector backbone to generate the R3-rep, which was sequenced and found to encode additional changes that included the L(2155)P substitution in the NS5A segment of the HCV polyprotein (compare R3-rep sequence with the R3 sequence in tables 2 and 3). Various quantities of *in vitro* transcribed R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of  $1.2 \times 10^6$  cfu/μg of RNA (panel C). Various quantities of R3-rep-5'G were also transfected resulting in a colony formation efficiency of  $2 \times 10^6$  cfu/μg of RNA (panel D).

**Figure 7** displays a typical RT-PCR amplification plot (left panel) and the graphical representation of Ct values versus known HCV RNA quantity in a standard curve (right panel). Each of the plotted curves in the left panel, graph the increment of fluorescence reporter signal (delta-Rn) versus PCR cycle number for a predetermined quantity of HCV replicon RNA. The Ct value is obtained by determining the point at which the fluorescence exceeds an arbitrary value (horizontal line). The right panel demonstrates the linear relationship between starting RNA copy number of the predetermined standards (large black dots) and the Ct value. Smaller dots are the Ct values of RNA samples (containing unknown quantity of HCV replicon RNA) from S22.3 cells treated with various concentrations of a specific inhibitor of HCV replication.

**Figure 8** shows the effect of increasing concentration of inhibitor A on HCV RNA replicon levels in Huh7 cells. S22.3 cells were grown in the presence of increasing concentrations of inhibitor A starting at 0.5nM and ranging to 1024nM. The inhibitor dose-response curve is the result of 11 concentrations from serial two-fold dilutions (1:1). One control well, without any inhibitor, was also included during the course of

the experiment. The cells were incubated for 4 days in a 5% CO<sub>2</sub> incubator at 37 °C. Total cellular RNA was extracted, quantified by optical density. HCV replicon RNA was evaluated by real time RT-PCR and plotted as genome equivalents/μg total RNA as a function of inhibitor concentration

5

### Definitions

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell

10 culture, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook *et al.* (1989) and Ausubel *et al.* (1994).

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art

15 and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission (1972).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

20 The term "DNA segment or molecule or sequence", is used herein, to refer to molecules comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). These segments, molecules or sequences can be found in nature or synthetically derived. When read in accordance with the genetic code, these sequences can encode a linear stretch or sequence of amino acids which can

25 be referred to as a polypeptide, protein, protein fragment and the like.

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.

30 A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific structural function that constitute the viral particles. "Structural proteins" defines the HCV proteins incorporated into the virus particles namely, core "C", E1, E2, and E2-p7.

"Non-structural proteins", defines the HCV proteins that are not comprised in viral

35 particles namely, NS2, NS3, NS4A, NS5A and NS5B.

"Restriction endonuclease or restriction enzyme" is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5 or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. An example of such an enzyme is *EcoRI*, which recognizes the base  
5 sequence G↓AATTC and cleaves a DNA molecule at this recognition site.

"Restriction fragments" are DNA molecules produced by the digestion of DNA with a restriction endonuclease. Any given genome or DNA segment can be digested by a particular restriction endonuclease into at least two discrete molecules of restriction fragments.

10 "Agarose gel electrophoresis" is an analytical method for fractionating polynucleotide molecules based on their size. The method is based on the fact that nucleic acid molecules migrate through a gel as through a sieve, whereby the smallest molecule has the greatest mobility and travels the farthest through the gel. The sieving characteristics of the gel retards the largest molecules such that, these have the  
15 least mobility. The fractionated polynucleotides can be visualized by staining the gel using methods well known in the art, nucleic acid hybridization or by tagging the fractionated molecules with a detectable label. All these methods are well known in the art, specific methods can be found in Ausubel *et al. (supra)*.

"Oligonucleotide or oligomer" is a molecule comprised of two or more  
20 deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size of the molecule will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. An oligonucleotide can be derived synthetically, by cloning or by amplification.

"Sequence amplification" is a method for generating large amounts of a target  
25 sequence. In general, one or more amplification primers are annealed to a nucleic acid sequence. Using appropriate enzymes, sequences found adjacent to, or in between the primers are amplified. An amplification method used herein is the polymerase chain reaction (PCR) and can be used in conjunction with the reverse-transcriptase (RT) to produce amplified DNA copies of specific RNA sequences.

30 "Amplification primer" refers to an oligonucleotide, capable of annealing to a RNA or DNA region adjacent to a target sequence and serving as the initiation primer for DNA synthesis under suitable conditions well known in the art. The synthesized primer extension product is complementary to the target sequence.

The term "domain" or "region" refers to a specific amino acid sequence that defines  
35 either a specific function or structure within a protein. As an example herein, is the

## 12

NS3 protease domain comprised within the HCV non-structural polyprotein.

The terms "plasmid" "vector" or "DNA construct" are commonly known in the art and refer to any genetic element, including, but not limited to, plasmid DNA, phage DNA, viral DNA and the like which can incorporate the oligonucleotide sequences, or  
5 sequences of the present invention and serve as DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The terminology "expression vector" defines a vector as described above but designed to enable the expression of an inserted sequence following transformation  
10 or transfection into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. Such expression control sequences will vary depending on whether the vector is designed to express the operably linked gene *in vitro* or *in vivo* in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements  
15 such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) or RNA, when such nucleic acid has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently  
20 linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting/transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, an example of a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome and is inherited by daughter cells through  
25 chromosome replication. A host cell or indicator cell can be transfected with RNA. A cell can be stably transfected with RNA if the RNA replicates and copies of the RNA segregate to daughter cells upon cell division. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA or RNA. Transfection methods are  
30 well known in the art (Sambrook *et al.*, 1989; Ausubel *et al.*, 1994). If the RNA encodes for a genetic marker that imparts an observable phenotype, such as antibiotic resistance, then the stable transfection of replicating RNA can be monitored by the acquisition of such phenotype by the host cell.

As used herein the term "transduction" refers to the transfer of a genetic marker to  
35 host cells by the stable transfection of a replicating RNA.

## 13

The nucleotide sequences and polypeptides useful to practice the invention include without being limited thereto, mutants, homologs, subtypes, quasi-species, alleles, and the like. It is understood that generally, the sequences of the present invention encode a polypeptide. It will be clear to a person skilled in the art that the polypeptide of the present invention and any variant, derivative or fragment thereof, is auto-processed to an active protease.

As used herein, the designation "variant " denotes in the context of this invention a sequence whether a nucleic acid or amino acid, a molecule that retains a biological activity (either functional or structural) that is substantially similar to that of the original sequence. This variant may be from the same or different species and may be a natural variant or be prepared synthetically. Such variants include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided the biological activity of the protein is conserved. The same applies to variants of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained.

The term "derivative" is intended to include any of the above described variants when comprising additional chemical moiety not normally a part of these molecules. These chemical moieties can have varying purposes including, improving a molecule's solubility, absorption, biological half life, decreasing toxicity and eliminating or decreasing undesirable side effects. Furthermore, these moieties can be used for the purpose of labeling, binding, or they may be comprised in fusion product(s). Different moieties capable of mediating the above described effects can be found in *Remington's The Science and Practice of Pharmacy* (1995).

Methodologies for coupling such moieties to a molecule are well known in the art. The term "fragment" refers to any segment of an identified DNA, RNA or amino acid sequence and/or any segment of any of the variants or derivatives described herein above that substantially retains its biological activity (functional or structural) as required by the present invention.

The terms "variant", "derivative", and "fragment" of the present invention refer herein to proteins or nucleic acid molecules which can be isolated/purified, synthesized chemically or produced through recombinant DNA technology. All these methods are well known in the art. As exemplified herein below, the nucleotide sequences and polypeptides used in the present invention can be modified, for example by *in vitro* mutagenesis.



As used herein, the term "HCV polyprotein coding region" means the portion of a hepatitis C virus that codes for the polyprotein open reading frame (ORF). This ORF may encode proteins that are the same or different than wild-type HCV proteins. The ORF may also encode only some of the functional protein encoded by wild-type polyprotein coding region. The protein encoded therein may also be from different isolates of HCV, and non-HCV protein may also be encoded therein.

As used herein, the abbreviation "NTR" used in the context of a polynucleotide molecule means a non-translated region. The term "UTR" means untranslated region. Both are used interchangeably.

#### Preferred embodiments

Particularly, the invention provides a HCV self-replicating polynucleotide molecule comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO.8).

According to the first embodiment of this invention, there is particularly provided a HCV polynucleotide construct comprising:

- a 5'-non translated region (NTR) comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein coding region; and
- a 3'-NTR region.

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

Alternatively, the first embodiment of the present invention is directed to HCV self-replicating polynucleotide molecule comprising a G2042C/R mutation.

## 15

According to the second embodiment, the present invention particularly provides a HCV polynucleotide construct comprising:

- 5       - a 5'-NTR region comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein region coding for a HCV polyprotein comprising a G(2042)C or a G(2042)R mutation; and
- a 3'-NTR region.

10      Preferably, the polynucleotide construct of the present invention is a DNA or RNA molecule. More preferably, the construct is a RNA molecule. Most preferably, the construct is a DNA molecule.

      More particularly, the first embodiment of this invention is directed to a RNA  
15      molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.

      Most particularly, the invention provides a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.

20      In a third embodiment, the invention also is directed to an expression vector comprising DNA forms of the above polynucleotide, operably linked with a promoter.

      Preferably, the promoter is selected from the group consisting of: T3, T7 and SP6.

25      According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or vector as described above. Particularly, the host cell is a eukaryotic cell line. More particularly, the eukaryotic cell line is a hepatic cell line. Most particularly, the hepatic cell line is Huh-7.

30      In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- a) incubating the host cell as described above under conditions suitable for RNA replication;
- 35      b) isolating the total cellular RNA from the cells; and

## 16

c) analyzing the RNA so as to measure the amount of HCV RNA replicated.

Preferably, the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the  
5 amount of HCV RNA replicated.

Alternatively in this fifth embodiment, the construct comprises a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.  
10

According to a preferred aspect of the sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- 15 a) carrying step a) as described in the above assay, in the presence or absence of the compound;  
b) isolating the total cellular RNA from the cells; and  
c) analyzing the RNA so as to measure the amount of HCV RNA replicated.  
d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,  
20 wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

Preferably, the cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.  
25

## EXAMPLES

### EXAMPLE 1

#### Replicon Constructs (APGK-12; Figure 1)

30 pET9a-EMCV was obtained by ligating an oligonucleotide linker  
5' gaattccagatggcgcgcccagatgttaaccagatccatggcacactctagagtactgtcgac 3' (SEQ ID NO.9) to pET-9a (Novagen) that was cut with EcoRI and Sall to form the vector pET-9a-mod. This linker contains the following restriction sites: EcoRI, Ascl, HpaI, NcoI, XbaI, Scal, Sall. The EMCV IRES was amplified by PCR from the vector pTM1 with  
35 primers

5' cggaatcgtaacagaccacaacggttcctc 3' (SEQ ID NO.10) and 5' ggcgtacccatggtattatcggttttca 3' (SEQ ID NO.11) and ligated into pET-9a-mod via EcoRI and NcoI to form pET-9a-EMCV.

- 5 The sequence of HCV NS2 to NS5B followed by the 3'UTR of HCV was obtained from the replicon construct I377/NS2-3' (Lohman *et al.*, 1999; accession number: AJ242651) and synthesized by Operon Technologies Inc. with a T to C change at the NcoI site in NS5B at nucleotide 8032. This sequence was released from an GenOp® vector (Operon Technologies) with NcoI and Scal and transferred into pET-  
10 9a-EMCV to form pET-9a-EMCV-NS2-5B-3'UTR.

pET-9a-HCV-neo was obtained by amplification of the HCV IRES from a HCV cDNA isolated from patient serum with primers

- 5' gcatatgaattctaatacgaactcactataggccagcccccgattg 3' (SEQ ID NO.12) containing a  
15 T7 promoter and primer

5' ggcgcgcccttgggtttcttggaggttaggattcgctcat 3' (SEQ ID NO.13) and amplification of the neomycin phosphotransferase gene from the vector pcDNA 3.1 (Invitrogen) with primers

- 5' aaaggcgcgatgattgaacaagatggattgcacgca 3' (SEQ ID NO.14) and 5'  
20 gcatatgttaactcagaagaactcgtaagaaggcgata 3' (SEQ ID NO.15). These two PCR fragments were mixed and amplified with primers  
5' gcatatgaattctaatacgaactcactataggccagcccccgattg 3' (SEQ ID NO.16) and  
5' gcatatgttaactcagaagaactcgtaagaaggcgata 3' (SEQ ID NO.15); cut with Eco RI and HpaI and transferred into pET-9a-mod to form pet-9a-HCV-neo. The EMCV-  
25 NS2-5B-3'UTR was released from pET-9a-EMCV-NS2-5B-3'UTR with HpaI and Scal and transferred into pet-9a-HCV-neo that was cut with HpaI to form pET-9a-APGK12. This insert was sequenced with specific successive primers using a ABI Prism® BigDye™ Terminator Cycle sequencing kit and analyzed on ABI Prism® 377 DNA Sequencer and is shown in SEQ ID NO 1.

30

#### **RNA *in vitro* transcription**

- pET-9a-APGK12 DNA was cut with Scal for expression of the full-length replicon or with BglII for expression of a truncated negative control RNA. DNA was analyzed on a 1% agarose gel and purified by Phenol/Chloroform extraction. RNA was produced  
35 using a T7 Ribomax® kit (Promega) followed by extraction with phenol/chloroform

## 18

and precipitation with 7.5 M LiCl<sub>2</sub>. RNA was treated with DNase I for 15 min to remove the DNA template and further purified with an RNeasy® column (Qiagen). RNA integrity was verified on a denaturing formaldehyde 1% agarose gel.

5 **EXAMPLE 2****Primary transfection of Huh7 cells and selection of replicon cell lines**

Human hepatoma Huh7 cells (Health Science Research Resources Bank, Osaka, Japan) were grown in 10% FBS/DMEM. Cells were grown to 70% confluency, trypsinized, washed with phosphate buffered saline (PBS) and adjusted to 1x10<sup>7</sup> cells/ml of PBS. 800 µl of cells were transferred into 0.4cm cuvettes and mixed with 15 µg of replicon RNA. Cells were electroporated using 960µF, 300 volts for ~18 msec and evenly distributed into two 15 cm tissue culture plates and incubated in a tissue culture incubator for 24 hours. The selection of first and second generation replicon cell lines was with 10% FBS/DMEM medium supplemented with 1mg/ml of G418. Cells were selected for 3-5 weeks until colonies were observed that were isolated and expanded.

Following the G418 selection and propagation of Huh-7 cells transfected with APGK12 (SEQ ID NO. 1) RNA, cells that formed a distinct colony were treated with trypsin and serially passed into larger culture flasks to establish cell lines. Approximately 10 X 10<sup>6</sup> cells were harvested from each cell line. The cells were lysed and the total cellular RNA extracted and purified as outlined in Qiagen RNAeasy® preparatory procedures. Figure 2 shows the analysis of 12 µg of total cellular RNA from various cell lines as analyzed on a Northern blot of a denaturing agarose-formaldehyde gel.

Figure 2A is a Northern blot (radioactively probed with HCV specific minus-strand RNA) that detects the presence of plus-strand replicon RNA. Lanes 1 and 2 are positive controls that contain 10<sup>9</sup> copies of *in vitro* transcribed APGK12 RNA. Lane 2 contains the *in vitro* transcribed RNA mixed with 12 µg of total cellular from naïve Huh-7 cells. Lane 3 is a negative control of total cellular RNA from untreated Huh-7 cells. Lanes 4 and 5 contain cellular RNA from the B1 and B3 G418 resistant cell lines that have DNA integrated copies of the neomycin phosphotransferase gene. Lane 6 contains total cellular RNA from a Huh-7 cell line, designated S22.3, that

harbors high copy number of HCV sub-genomic replicon RNA as detected by the positive signal in the 8 kilo-base range. Other cell lines have no detectable replicon RNA. Figure 2B is a Northern blot of a duplicate of the gel presented in 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA (lanes 1 and 2 are positive control lanes that contain  $10^9$  copies of full length genomic HCV minus strand RNA); only lane 6, which contains 12  $\mu$ g of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8 – 9 kilobases. An quantitative estimation of RNA copy number, based on phosphorimager scanning of the Northern blots, is approximately  $6 \times 10^7$  copies of plus-strand/ $\mu$ g of total RNA, and  $6 \times 10^6$  copies of minus strand/ $\mu$ g of total RNA. The presence of the plus-strand and minus-strand intermediate confirms that the HCV sub-genomic RNA is actively replicating in the S22.3 cell line.

### EXAMPLE 3

#### **S22.3 cell line constitutively expresses HCV non-structural proteins.**

HCV non-structural protein expression was examined in the S22.3 cell line. Figure 3 displays the result of indirect immunofluorescence that detects the HCV NS4A protein in the S22.3 cell line and not in the replicon negative B1 cell line (a G418 resistant Huh-7 cell line). Indirect immunofluorescence was performed on cells that were cultured and fixed (with 4% paraformaldehyde) onto Lab-tek chamber slides. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes followed by a 1 hour treatment with 5% milk powder dissolved in phosphate-buffered saline (PBS). A rabbit serum containing polyclonal antibody raised against a peptide spanning the HCV NS4A region was the primary antibody used in detection. Following a 2 hour incubation with the primary antibody, cells were washed with PBS and a secondary goat anti-rabbit antibody conjugated with red-fluor Alexa® 594 (Molecular Probes) was added to cells for 3 hours. Unbound secondary antibody was removed with PBS washes and cells were sealed with a cover slip. Figure 3 (top panels) shows the results of immunofluorescence as detected by a microscope with specific fluorescent filtering; the bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have

background level of staining. A small proportion of S22.3 cells express high levels of intensely stained HCV NS4A.

Expression of the proteins encoded by the bi-cistronic replicon RNA was also  
5 examined on Western-blots following SDS-PAGE separation of total proteins  
extracted from: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1,  
and (iii) the S22.3 cell line. Figure 4 panels A, B, and C, demonstrate the results of  
western blots probed with rabbit polyclonal antisera specific for neomycin  
phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization  
10 was achieved through autoradiographic detection of a chemiluminescent reactive  
secondary HRP-conjugated goat anti-rabbit antibody. Figure 4 panel A shows that  
the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than B1  
cells (which contain an integrated DNA copy of the *npt* gene) and that the naïve  
Huh-7 cell line does not produce the NPT protein. Figure 4 panels B and C show  
15 that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins,  
respectively. The western blots demonstrate that the S22.3 cell line, which harbors  
actively replicating HCV sub-genomic replicon RNA, maintains replication of the  
RNA through the high level expression of the HCV non-structural proteins.

#### 20 EXAMPLE 4

##### Sequence determination of adapted replicons

Total RNA was extracted from replicon containing Huh7 cells using a RNeasy Kit  
(Qiagen). Replicon RNA was reverse transcribed and amplified by PCR using a  
25 OneStep RT-PCR kit (Qiagen) and HCV specific primers (as selected from the full-  
length sequence disclosed in WO 00/66623). Ten distinct RT-PCR products, that  
covered the entire bi-cistronic replicon in a staggered fashion, were amplified using  
oligonucleotide primers. The PCR fragments were sequenced directly with ABI  
Prism® BigDye™ Terminator Cycle PCR Sequencing and analyzed on ABI Prism®  
30 377 DNA Sequencer. To analyze the sequence of the HCV replicon 3' and 5' ends a  
RNA ligation/RT-PCR procedure described in Kolykhalov *et al.* 1996 was followed.  
The nucleotide sequence of S22.3 is presented as **SEQ ID NO. 2**.

## 21

**EXAMPLE 5.****Serial Passage of HCV Replicon RNA**

The total cellular RNA from the S22.3 cell line was prepared as described above.

5 HCV Replicon RNA copy number was determined by Taqman® RT-PCR analysis and 20 µg of total S22.3 cellular RNA (containing  $1 \times 10^9$  copies of HCV RNA) was transfected by electroporation into  $8 \times 10^6$  naïve Huh-7 cells. Transfected cells were subsequently cultured in 10 cm tissue culture plates containing DMEM supplemented with 10% fetal calf serum (10% FCS). Media was changed to DMEM

10 (10% FCS) supplemented with 1 mg/ml G418 24 hours after transfection and then changed every three days. Twenty-three visible colonies formed three to four weeks post-transfection and G418 selection. G418 resistant colonies were expanded into second generation cell lines that represent the first cell lines harboring serially passaged HCV Replicon RNA. Three of these cell lines: R3, R7, and R16 were the

15 subject of further analyses. First, the efficiency of transduction by each of the adapted replicons was determined by electroporation of the total cellular RNA (extracted from the R3, R7 and R16) into naïve Huh-7 cells; following electroporation, the transduction efficiency was determined as described above, by counting the visible G418 resistant colonies that arose following 3 to 5 weeks of

20 G418 selection (Table 1). Second, the sequence of the serially passed adapted replicons was determined from the total cellular RNA that was extracted from each of the R3, R7 and R16 replicon cell lines as described in example 4 (SEQ ID NO. 4, 5, 6). Using the pAPGK12 as a reference sequence (SEQ ID NO. 1), the nucleotide changes that were selected in HCV segment of the adapted replicons are presented

25 in Figure 5A. Some of these nucleotide changes are silent and do not change the encoded amino acid whereas others result in an amino acid substitution. Figure 5B summarizes the amino acid changes encoded by the adapted replicons with the amino acid sequence of pAPGK12 as the reference. It is important to note that the reference sequence APGK-12 (SEQ ID NO.1) contains an extra G at the 5'-terminal (5'-GG) that is not maintained in the replicating RNA of the established cell lines.

30 Also noteworthy is that, in addition to G->A at nucleotide 1, there is also an adapted mutation G->C/R at amino acid 2042 (shown as amino acid 1233 in the sequence listing since a.a. 810 of NS2 is numbered as a.a. 1 in SEQ ID) that can be found in all clones analyzed.



## 22

**TABLE 1**  
Transfection of Huh-7 cells

	<u>RNA</u>	<u>Copies of Replicon</u>	<u># Colonies</u>	<u>SEQ ID</u>
5	5 ng APKG12 replicon in 20µg total Huh-7 RNA	$1.2 \times 10^9$	0	
10	15 µg APKG12 replicon RNA	$3 \times 10^{12}$	1 (S22.3)	1
	20µg total: S22.3 cellular RNA	$3 \times 10^9$	23 (3 clones analyzed)	2
15	<b>R3 cellular RNA</b>	$1 \times 10^9$	<b>200</b>	<b>4</b>
	R7 cellular RNA	$1 \times 10^9$	20	5
	<b>R16 cellular RNA</b>	$3 \times 10^8$	<b>100</b>	<b>6</b>
	<b>cloned R3rep RNA</b>	$2.3 \times 10^8$	<b>2000</b>	<b>7</b>

20 **EXAMPLE 6**

**Construction of APGK12 with 5' G-> A substitution (APGK12-5'A, SEQ ID NO.24)**

The pAPGK12 DNA was modified to change the first nucleotide in the sequence to replace the 5'GG with a 5'A. The change in the pAPGK12 was introduced by replacing an *EcoRI*/*Agel* portion of the sequence with a PCR-generated *EcoRI*/*Agel* fragment that includes the mutation. The oligonucleotides used for the amplification were (SEQ ID. NO. 20): 5'-GTG GAC GAA TTC TAA TAC GAC TCA CTA TAA CCA GCC CCC GAT TGG-3' and (SEQ ID. NO. 21): 5'-GGA ACG CCC GTC GTG GCC AGC CAC GAT-3' and generated a 195 bp DNA fragment that was then digested with *EcoRI* and *Agel*. The resulting 178 bp restriction fragment was used to replace the *EcoRI* / *Agel* fragment in pAPGK12 to generate the pAPGK12-5'A plasmid.

**EXAMPLE 7**

**CDNA CLONING OF THE R3-REPLICON (R3REP).**

The cDNA clone of the R3 replicon was produced by RT-PCR of RNA extracted from the R3 cell line. The following two oligonucleotides were used: (SEQ ID. NO. 22): 5'-GTC GTC TTC TCT GAC ATG GAG AC-3' and (SEQ ID. NO. 23): 5'-GAG TTG

## 23

CTC AGT GGA TTG ATG GGC AGC-3'. The ~4400nt PCR fragment, starting within the NS2 coding region and extending to the 5'-end of the NS5B coding region, was cloned into the plasmid pCR3.1 by TA cloning (Invitrogen). The *SacII* / *XhoI* portion of this R3 sequence was then used to replace the *SacII* / *XhoI* fragment present in the pAPGK12 and the pAPGK12-5'A described above. Consequently, two R3 cDNA sequences were generated: (I) R3-Rep-5'G with an initiating 5'G (SEQ ID NO.7), and R3-Rep-5'A (SEQ ID NO.25) with an initiating 5'A. Sequencing of the R3 rep cDNA identified unique nucleotide changes that differ from the original pAPGK12 sequence (see Figure 5A); some of these changes are silent and do not change the encoded amino acid, whereas others do result in an amino acid change (see Figure 5B). The differences between R3 and the R3-rep reflect the isolation of a unique R3-rep cDNA clone encoding nucleotide changes that were not observed from the sequencing of the total RNA extracted from the R3 cell line.

15 **EXAMPLE 8****Efficiency of colony formation with modified constructs**

RNA from pAPGK12, pAPGK12-5'A, pR3-Rep and pR3-Rep-5'A was generated by *in vitro* transcription using the T7 Ribomax® kit (Promega) as described in example 1 above. The reactions containing the pAPGK12-5'A and pR3-Rep-5'A templates were scaled-up 10-fold due to the limitation of commercial RNA polymerase in initiating transcripts with 5'-A. The full length RNAs and control truncated RNA for each clone were introduced into  $8 \times 10^6$  naïve Huh-7 cells by electroporation as described in example 2. Replicon RNA was supplemented with total cellular Huh-7 carrier RNA to achieve a final 15-20µg quantity. The cells were then cultured in DMEM medium supplemented with 10% fetal calf serum and 0.25 mg/ml G418 in two 150 mm plates. The lower concentration of G418 was sufficient to isolate and select replicon containing cell lines as none of the transfectants with the control truncated RNA produced any resistant colonies. In contrast, *in vitro* transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in Figure 6 panels A and B) following selection with G418. Various quantities (ranging from 0.1 ng to 1 µg) of the R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of  $1.2 \times 10^6$  cfu/µg of RNA (Figure 6 panel C depicts transfection with 1 µg of RNA). Various quantities (ranging from 0.1 ng to 1 µg) of R3-rep [5'G] were similarly transfected resulting in a colony formation efficiency of  $2 \times 10^6$  cfu/µg of RNA (Figure 6 panel D

depicts colony formation with 1 µg of RNA). Note that, shown for the first time, HCV subgenomic replicons replicate as efficiently with a 5' A nucleotide in place of the 5'G. APGK12 with a 5'A or 5'G RNA have similar transduction efficiencies. Similarly, R3-Rep RNAs with either the 5'A or 5'G both display the markedly increased transduction efficiency. Notably, the adaptive mutants within the HCV non-structural segment encoded by the R3-Rep provides for a substantial increase in transduction efficiency as depicted by the dramatic increase in colony forming units per µg of transfected RNA.

#### 10 EXAMPLE 9

##### Quantification of HCV Replicon RNA Levels in Cell lines

S22.3 cells, or cell lines harboring other adapted replicons, were seeded in DMEM supplemented with 10% FBS, PenStrep and 1 µg/mL Geneticin. At the end of the incubation period the replicon copy number is evaluated by real-time RT-PCR with the ABI Prism 7700 Sequence Detection System. The TAQMAN® EZ RT-PCR kit provides a system for the detection and analysis of HCV RNA (as first demonstrated by Martell *et al.* 1999 J. Clin. Microbiol. 37: 327-332). Direct detection of the reverse transcription polymerase chain reaction (RT-PCR) product with no downstream processing is accomplished by monitoring the increase in fluorescence of a dye-labeled DNA probe (Figure 6). The nucleotide sequence of both primers (adapted from Ruster, B. Zeuzem, S. and Roth, W.K., 1995. Analytical Biochemistry 224:597-600) and probe (adapted from Hohne, M., Roeske, H. and Schreier, E. 1998, Poster Presentation: P297 at the Fifth International Meeting on Hepatitis C Virus and Related Viruses Molecular Virology and Pathogenesis, Venezia-Lido Italy, June 25-28, 1998) located in the 5'-region of the HCV genome are the following:

HCV Forward primer:

5' ACG CAG AAA GCG TCT AGC CAT GGC GTT AGT 3' (SEQ ID NO.17)

30

HCV Reverse primer:

5' TCC CGG GGC ACT CGC AAG CAC CCT ATC AGG 3' (SEQ ID NO.18)

HCV Probe:

5' FAM-TGG TCT GCG GAA CGG GTG AGT ACA CC-TAMRA 3' (SEQ ID NO.19)

- 5 FAM: Fluorescence reporter dye.  
TAMRA: Quencher dye.

Using The TAQMAN® EZ RT-PCR kit, the following reaction was set up:

Component	Volume per sample ( $\mu$ L)	Final Concentration
RNase-Free Water	16	-
5X Taqman EZ Buffer	10	1X
Manganese Acetate 25mM	6	3mM
dATP 10mM	1.5	300 $\mu$ M
dCTP 10mM	1.5	300 $\mu$ M
dGTP 10mM	1.5	300 $\mu$ M
dUTP 20mM	1.5	300 $\mu$ M
HCV Forward Primer 10 $\mu$ M	1	200nM
HCV Reverse Primer 10 $\mu$ M	1	200nM
HCV Probe 5 $\mu$ M	2	200nM
rTth DNA Polymerase 2.5U/ $\mu$ L	2	0.1U/ $\mu$ L
AmpErase UNG 1U/ $\mu$ L	0.5	0.01U/ $\mu$ L
<b>Total Mix</b>	<b>45</b>	<b>-</b>

10

To this reaction mix, 5 $\mu$ L of total RNA extracted from S22.3 cells diluted at 10ng/ $\mu$ L was added, for a total of 50ng of RNA per reaction. The replicon copy number was evaluated with a standard curve made from known amounts of replicon copies (supplemented with 50ng of wild type Huh-7 RNA) and assayed in an identical reaction mix (Figure 7).

15

Thermal cycler parameters used for the RT-PCR reaction on the ABI Prism 7700 Sequence Detection System were optimized for HCV detection:

Cycle	Temperature (°C)	Time (Minutes)	Repeat	Reaction
Hold	50	2		Initial Step
Hold	60	30		Reverse Transcription
Hold	95	5		UNG Deactivation
Cycle	95	0:15	2	Melt
	60	1		Anneal/Extend
Cycle	90	0:15	40	Melt
	60	1		Anneal/Extend

Quantification is based on the threshold cycle, where the amplification plot crosses a defined fluorescence threshold. Comparison of the threshold cycles provides a highly sensitive measure of relative template concentration in different samples.

- 5 Monitoring during early cycles, when PCR fidelity is at its highest, provides precise data for accurate quantification. The relative template concentration can be converted to RNA copy numbers by employing a standard curve of HCV RNA with known copy number (Figure 7).

#### 10 EXAMPLE 10

**A specific HCV NS3 protease anti-viral compound inhibits replication of the HCV replicon in S22.3 cell lines.**

- In order to determine the effect of a specific HCV NS3 protease anti-viral compound on replicon levels in S22.3 cells, the cells were seeded in 24 Well Cell Culture Cluster at  $5 \times 10^4$  cells per well in 500 $\mu$ L of DMEM complemented with 10% FBS, PenStrep and 1 $\mu$ g/mL Geneticin. Cells were incubated until compound addition in a 5% CO<sub>2</sub> incubator at 37 °C. The dose-response curve of the inhibitor displayed 11 concentrations resulting from serial two-fold dilutions (1:1). The starting concentration of compound A was 100nM. One control well (without any compound) was also included in the course of the experiment. The 24 well plates were incubated for 4 days in a 5% CO<sub>2</sub> incubator at 37 °C. Following a 4 day incubation period, the cells were washed once with PBS and RNA was extracted with the RNeasy® Mini Kit and Qias shredder® from Qiagen. RNA from each well was eluted in 50 $\mu$ L of H<sub>2</sub>O. The RNA was quantified by optical density at 260nm on a Cary 1E UV-Visible Spectrophotometer. 50 ng of RNA from each well was used to quantify the HCV replicon RNA copy number as detailed in Example 6. The level of inhibition (% inhibition) of each well containing inhibitor was calculated with the following

equation (CN = HCV Replicon copy number):

$$\% \cdot inhibition = \left( \frac{CN \cdot control - CN \cdot well}{CN \cdot control} \right) * 100$$

- 5 The calculated % inhibition values were then used to determine  $IC_{50}$ , slope factor ( $n$ ) and maximum inhibition ( $I_{max}$ ) by the non-linear regression routine NLIN procedure of SAS using the following equation:

$$\% \cdot inhibition = \frac{I_{max} \times [inhibitor]^n}{[inhibitor]^n + IC_{50}^n}$$

10

Compound A was tested in the assay at least 4 times. The  $IC_{50}$  curves were analyzed individually by the SAS nonlinear regression analysis. Figure 8 shows a typical curve and Table 2 shows the individual and average  $IC_{50}$  values of compound A. The average  $IC_{50}$  of compound A in the replication assay was 1.1nM.

15

**TABLE 2**

$IC_{50}$  of compound A in the S22.3 Cell line Replicon Assay.

Compound	$IC_{50}$ (nM)	Average $IC_{50}$ (nM)
<b>A</b>	1.2	
	1.2	
	1.0	
	0.9	
		<b>1.1 ± 0.2</b>

## 20 DISCUSSION

The reproducible and robust *ex vivo* propagation of hepatitis C virus, to levels required for the accurate testing of potential anti-viral compounds, has not been achieved with any system. As an alternative approach to studying the molecular mechanisms of hepatitis C virus RNA replication, selectable self-replicating bi-

25 cistronic RNAs were developed (Lohman *et al.*, 1999, Science 285:110-113; Bartenschlager CA 2,303,526). Minimally, these replicons encode for some or all of

the non-structural proteins and also carry a selectable marker such as the neomycin phosphotransferase. Though intracellular steady-state levels of these sub-genomic replicon RNAs among the selected clones is moderate to high, the frequency of generating G418-resistant colonies upon transfection of the consensus RNA described by Lohman *et al.* or Bartenschlager is very low. Less than 100 colonies are generated when 8 million cells are transfected with 1 µg of *in vitro* transcribed bicistronic replicon RNA. A low efficiency of colony formation was first noted by Lohmann *et al.* (1999 *et al.*, Science 285:110-113). Since then, Lohmann *et al.* (2001), Blight *et al.* (2000), and Guo *et al.* (2001), have isolated sub-genomic RNAs with markedly improved efficiencies in the colony formation assay. Lohmann *et al.*, 1999 originally reported that selection of sub genomic replicons may not involve the selection of adaptive mutants as serially passaged RNA did not demonstrate an improved transfection efficiency. Nevertheless, in an effort to characterize the function and fitness of replicating HCV RNA, we serially passaged the replicon RNA that was isolated from the first selected cell-line. Notably, a significant increase in colony forming efficiency was obtained from this experiment, even though the quantity of replicon RNA was orders of magnitude lower than originally used to transfect the *in vitro* transcribed RNA. Furthermore, a second round serial passage of replicon RNA from this first generation clone into naive Huh-7 cells provided for yet another increase in colony formation efficiency (Table 1).

Our analysis of replicating HCV RNAs identified several adaptive mutations that enhance the efficiency of colony formation by up to 4 orders of magnitude. Adaptive mutations were found in many non-structural proteins, as well as in the 5' non-translated region. The substitution of the 5'-GG doublet for a 5'-A as the inaugurating nucleotide of the HCV 5'-UTR is a variant of the HCV genome that has not been previously described, despite the sequencing of innumerable genotypes and subtypes from across the world. Our original replicon that carried a 5'-GG evolved to variants with either a single 5'-A or 5'-G, both of which showed equal transduction efficiency. We describe here the first report of a HCV genome that can tolerate and stably maintain a 5'A extremity. Moreover, we were successful in re-introducing this defined single nucleotide substitution into our cDNA clone and generate *in vitro* transcribed RNA harboring such an extremity to confirm that a 5'A functions as efficiently as a 5'G.

We have identified adaptive amino acid substitutions in the HCV non-structural proteins NS3, NS4A and NS5A in the R3 replicon, and a substitution in NS5B in the R7 clone (see Figure 5B). These mutations, particularly the combination defined by the R3-rep (SEQ ID NO. 7), when reconstituted into a cDNA clone and transcribed onto a RNA replicon, result in a significantly enhanced transduction efficiency of up to 20,000 fold from the original wild type APGK12 replicon RNA. However, the steady state levels of intracellular replicon RNA were comparable from each of the different isolated clones. This result suggests that the increase in replication efficiency by the adaptive mutations does not result in higher stable intracellular RNA levels due to higher RNA replication, but rather confers increased permissivity for establishing the replicon in a greater number of Huh7 cells. Such a phenotype may be manifested transiently, through an initial increase of the amount of *de novo* replication, that is required to surpass a defined threshold to establish persistently replicating RNAs within a population of dividing cells.

Recently three other groups also identified other distinct adaptive mutants. Lohmann *et al.* (2000) reported enhanced transduction efficiencies of up to 10,000 fold with mutations in NS3, NS4B, NS5A and NS5B. Blight *et al.* (2000) reported an augmentation of transduction efficiencies up to 20,000 fold with a single mutation in NS5A whereas Guo *et al.* (2001) reported increases in transduction efficiencies of 5,000-10,000 fold with a deletion of a single amino acid in NS5A. The amino acid substitutions that we describe here have not previously been identified as adaptive mutants that enhance the efficiency of RNA transfection and/or replication. One exception is the mutation of E1202G in NS3 that we found in both the R7 and R16 replicons. This adaptation was previously described by Guo *et al.* (2001) and Krieger *et al.* (2001). All other adaptive mutations, without exception, described herein are unpublished.

The development of selectable subgenomic HCV replicons has provided for potential avenues of exploration on HCV RNA replication, persistence, and pathogenesis in cultured cells. However, the low transduction efficiency with the HCV RNA-containing replicons as originally described (Lohmann *et al.*, 1999) showed that it was not a practical system for reverse genetics studies. The adaptive mutants described herein overcome the low transduction efficiency. In light of the recent descriptions of adaptive mutants by other groups, we note that adaptation can be



achieved by distinct mutations in different HCV NS proteins, although the level of adaptation can vary drastically. The replicons encoding adaptive mutants that are described herein are ideally suited for reverse genetic studies to identify novel HCV targets or host cell targets that may modulate HCV RNA replication or HCV replicon  
5 RNA colony formation. The adapted and highly efficient replicons are suitable tools for characterizing subtle genotypic or phenotypic changes that affect an easily quantifiable transduction efficiency.

Lastly, we have used our adapted HCV sub genomic replicon cell-line to  
10 demonstrate the proficient inhibition of HCV RNA replication by a specific small molecule inhibitor of the HCV NS3 protease. This is the first demonstration that an antiviral, designed to specifically inhibit one of the HCV non-structural proteins, inhibits HCV RNA replication in cell culture. Moreover, this compound and our S22.3 cell line validate the proposal that RNA replication is directed by the HCV non-  
15 structural proteins NS3 to NS5B. The assay that we have described and validated will be extremely useful in characterizing other inhibitors of HCV non-structural protein function in cell culture in a high throughput fashion.

All references found throughout the present disclosure are herein incorporated by  
20 reference whether they be found in the following list or not.

#### References

- Ago *et al.* 1999, Structure 7: 1417-1426  
Ausubel *et al.*, 1994, Current Protocols in Molecular Biology, Wiley, New York.  
25 Bartenschlager, R. *et al.*, 1993, J. Virol., 67, 3835-3844.  
Behrens *et al.*, 1998, J. Virol. 72, 2364  
Bressanelli *et al.* 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039  
Blight *et al.* 2000, Science 290: 1972-1974  
Cho *et al.*, 1998, J. Biol. Chem., 273, 15045  
30 Dash *et al.*, 1997, Am. J. Pathol., 151, 363 – 373  
Fournier *et al.* 1998, J. Gen. Virol. 79, 2376  
Gale Jr. *et al.* 1997 Virology 230, 217  
Grakoui, A. *et al.*, 1993(a), J. Virol. 67, 1385-1395.  
Grakoui A, *et al.*, 1993(b), Proc Natl Acad Sci USA, 90, 10583-7  
35 Guo *et al.* (2001) J. Virol. 8516-8523

- Hijikata, M. *et al.*, **1991**, Proc. Natl. Acad. Sci. USA. **88**, 5547-5551.  
Hijikata, M. *et al.*, **1993**, J. Virol. **67**, 4665-4675.  
Ikeda *et al.* **1998**, Virus Res. **56**, 157  
Ito *et al.* **1996**, J. Gen. Virol. **77**, 1043 – 1054
- 5 IUPAC-IUB Biochemical Nomenclature Commission, **1972**, Biochemistry, **11**, 1726-1732.
- Kolykhalov *et al.* **1996** J. of Virology, **7**, p. 3363-3371  
Khromykh *et al.*, **1997**, J. Virol. **71**, 1497  
Kim, D.W. *et al.*, **1995**, Biochem. Biophys. Res. Comm., **215**, 160-166.
- 10 Kim *et al.*, **1996**, Cell, **87**, 343;  
Kim *et al.*, **1998**, Structure, **6**, 89  
Kim *et al.*, **1999**, Arch. Virol, **144**, 329-343.  
Krieger *et al.* (**2001**) J. Virol. 4614-4624  
Kwong AD. *et al.*, **1998**, Antiviral Res., **40**, 1-18
- 15 Lanford *et al.* **1994**, Virology **202**, 606  
Lesburg *et al.* **1999**, Nat. Struct. Biol. **6**: 937-943  
Lohman *et al.* **1999**, Science **285**: 110-113  
Lohman *et al.* (**2001**) J. Virol. 1437-1449  
Love, R.A. *et al.*, **1996**, Cell, **87**, 331-342
- 20 Martell *et al.* **1999** J. Clin. Microbiol. **37**: 327-332  
Mizutani *et al.* **1996**, J. Virol. **70**, 7219 – 7223  
Moser *et al.*, **1998**, J. Virol. **72**, 5318  
Reed *et al.*, **1997**, J. Virol. **71**, 7187  
Sambrook *et al.*, **1989**, Molecular Cloning – A Laboratory Manual, Cold Spring
- 25 Harbor Labs.  
Shimizu *et al.* **1993**, PNAS, USA, **90**, 6037 – 6041  
Yanagi *et al.*, **1999**, Proc. Natl. Acad. Sci. USA, **96**, 2291-95  
Yao *et al.*, **1997**, Nature Structural Biology, **4**, 463  
Yem *et al.*, **1998**, Protein Science, **7**, 837
- 30 Yoo *et al.* **1995**, J. Virol., **69**, 32 – 38

## CLAIMS

1. A HCV polynucleotide molecule comprising a 5'-non translated region (NTR) wherein guanine at position 1 is substituted for adenine.
2. A HCV self-replicating polynucleotide comprising:
  - a 5'-NTR consisting of ACCAGC (SEQ ID NO. 8);
  - a HCV polyprotein region coding for a HCV polyprotein; and
  - a 3'-NTR region.
3. The HCV polynucleotide according to claim 2, wherein said polyprotein comprises one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.
4. The HCV polynucleotide encoding a polyprotein comprising one or more of the amino acid substitution as defined in claim 3, and further comprising the amino acid substitution E(1202)G.
5. The HCV polynucleotide according to claim 3, wherein said substitution is a G2042C or a G2042R mutation.
6. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: K(1691)R; and G(2042)C.
7. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
8. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
9. The HCV polynucleotide according to claim 3, wherein said substitution is selected

## 33

from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.

10. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
11. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
12. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
13. An expression vector comprising a DNA form of the polynucleotide according to claim 2, operably linked to a promoter.
14. A host cell transfected with the self-replicating polynucleotide molecule according to claim 2.
15. A host cell according to claim 14, wherein the host cell is a eukaryotic cell line.
16. A host cell according to claim 15, wherein said eukaryotic cell line is a hepatic cell line.
17. A host cell according to claim 16, wherein said hepatic cell line is Huh-7.
18. A RNA replication assay comprising the steps of:
  - a) incubating the host cell according to claim 14 under conditions suitable for RNA replication;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
19. The assay according to claim 18, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

20. The assay according to claim 18, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
21. A method for testing a compound for inhibiting HCV replication, including the steps of:
- a) carrying step a) according to claim 18, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,
- wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.
22. The method according to claim 21, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.
23. A HCV polynucleotide molecule comprising:
- a 5'-NTR region;
  - a HCV polyprotein region coding for a HCV polyprotein comprising one or more amino acid substitution selected from the group consisting of:  
R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A;  
G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T; and
  - a 3'-NTR region.
24. The HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as defined in claim 24, further comprising the amino acid substitution E(1202)G.
25. The polynucleotide according to claim 24, wherein said substitution is a G2042C or a G2042R mutation.
26. The HCV polynucleotide according to claim 24, wherein said substitution is selected

## 35

from the group consisting of: K(1691)R; and G(2042)C.

27. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
28. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
29. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
30. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
31. The HCV polynucleotide according to claim 24, wherein said molecule is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
32. The HCV polynucleotide according to claim 24, wherein said molecule is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
33. An expression vector comprising a DNA form of the polynucleotide according to claim 24, operably linked to a promoter.
34. A host cell transfected with the self-replicating polynucleotide according to claim 24.
35. A host cell according to claim 34, wherein the host cell is a eukaryotic cell line.
36. A host cell according to claim 35, wherein said eukaryotic cell line is a hepatic cell line.
37. A host cell according to claim 36, wherein said hepatic cell line is Huh-7.

- 38.** A RNA replication assay comprising the steps of:
- incubating the host cell according to claim 34 under conditions suitable for RNA replication;
  - isolating the total cellular RNA from the cells; and
  - analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 39.** The assay according to claim 38, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.
- 40.** The assay according to claim 38, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- 41.** A method for testing a compound for inhibiting HCV replication, including the steps of:
- a) carrying step a) according to claim 38, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,
- wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.
- 42.** The method according to claim 41, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

FIGURE 1

APGK12 (SEQ ID NO 1) Replicon RNA compared to I377/NS2-3' Replicon RNA

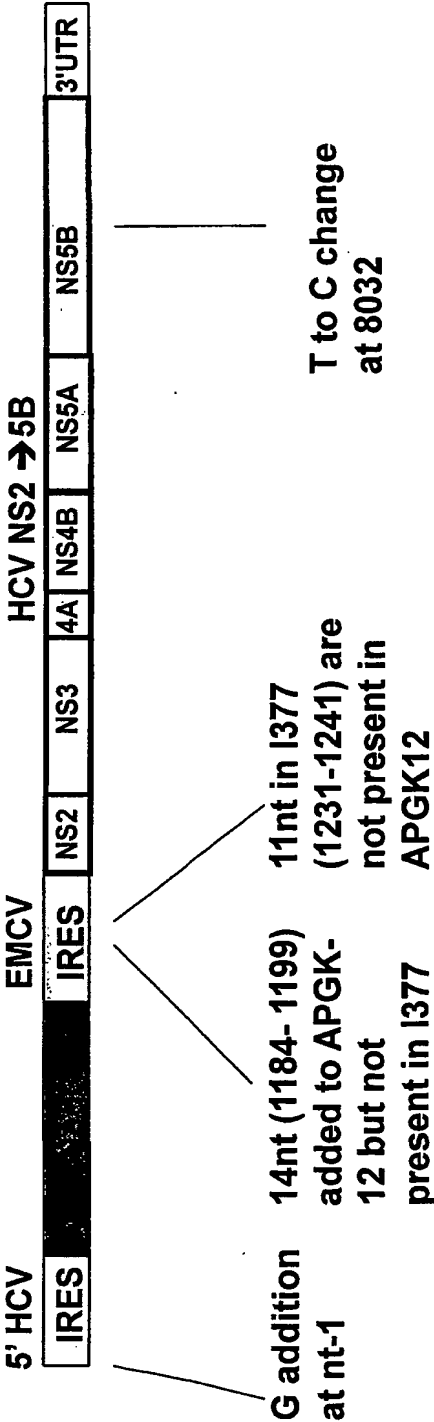
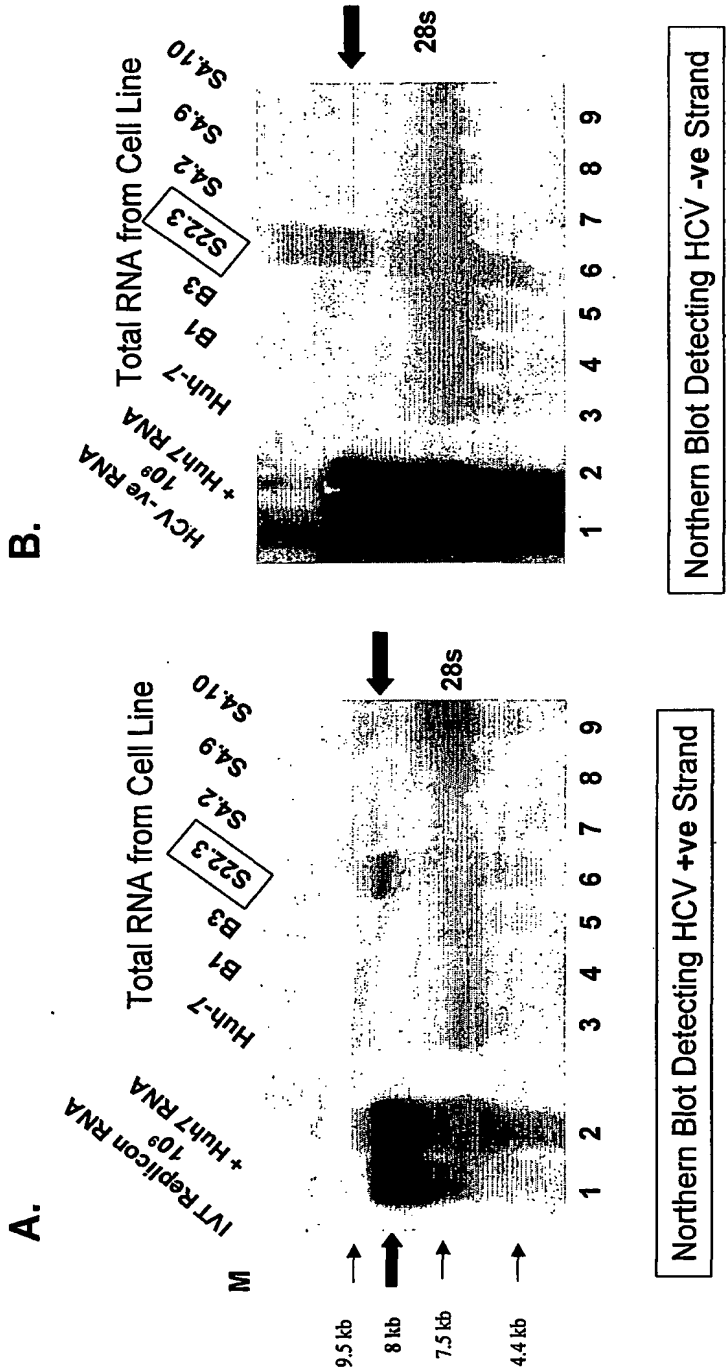




FIGURE 2



3 / 9

FIGURE 3

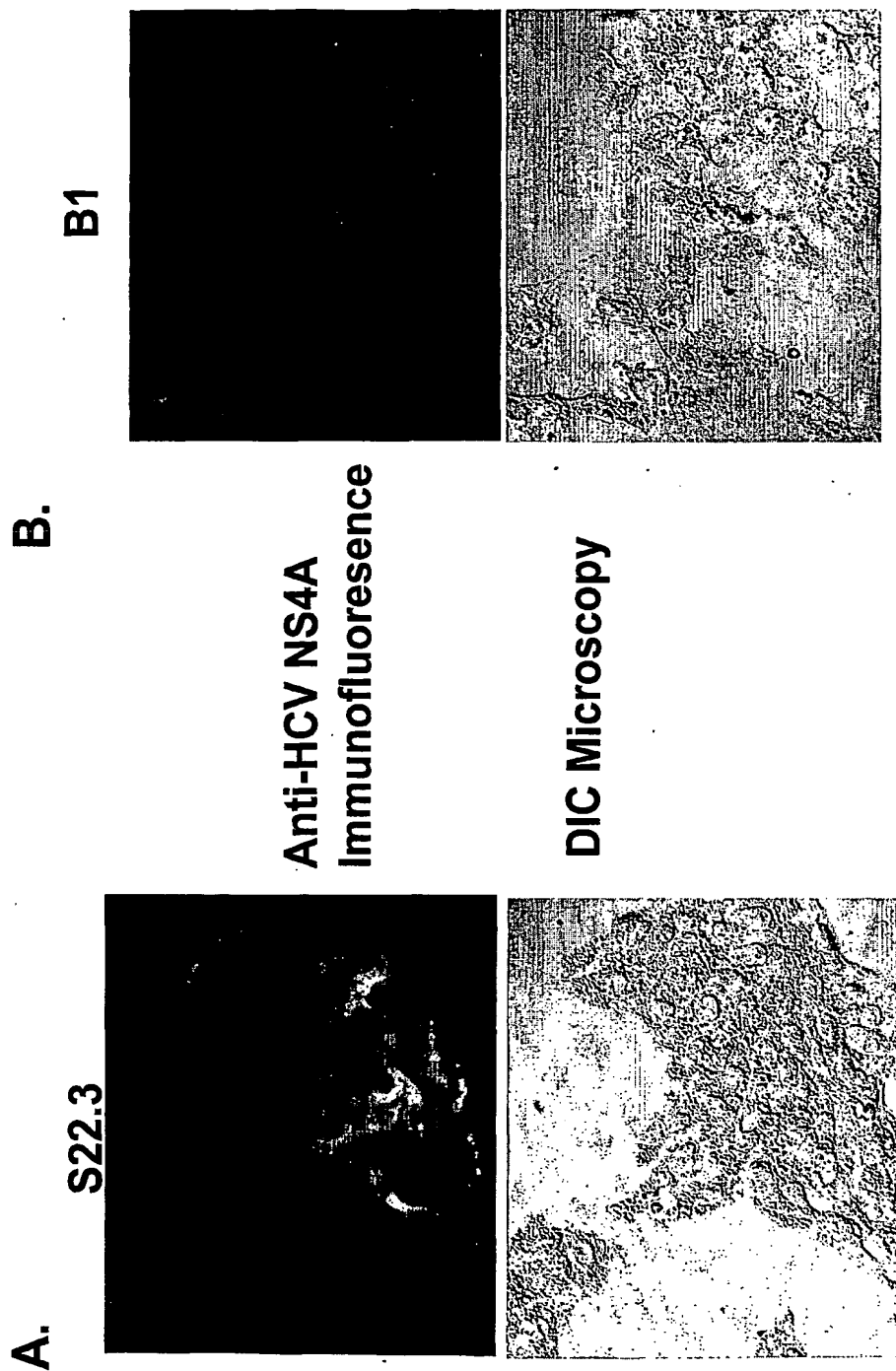
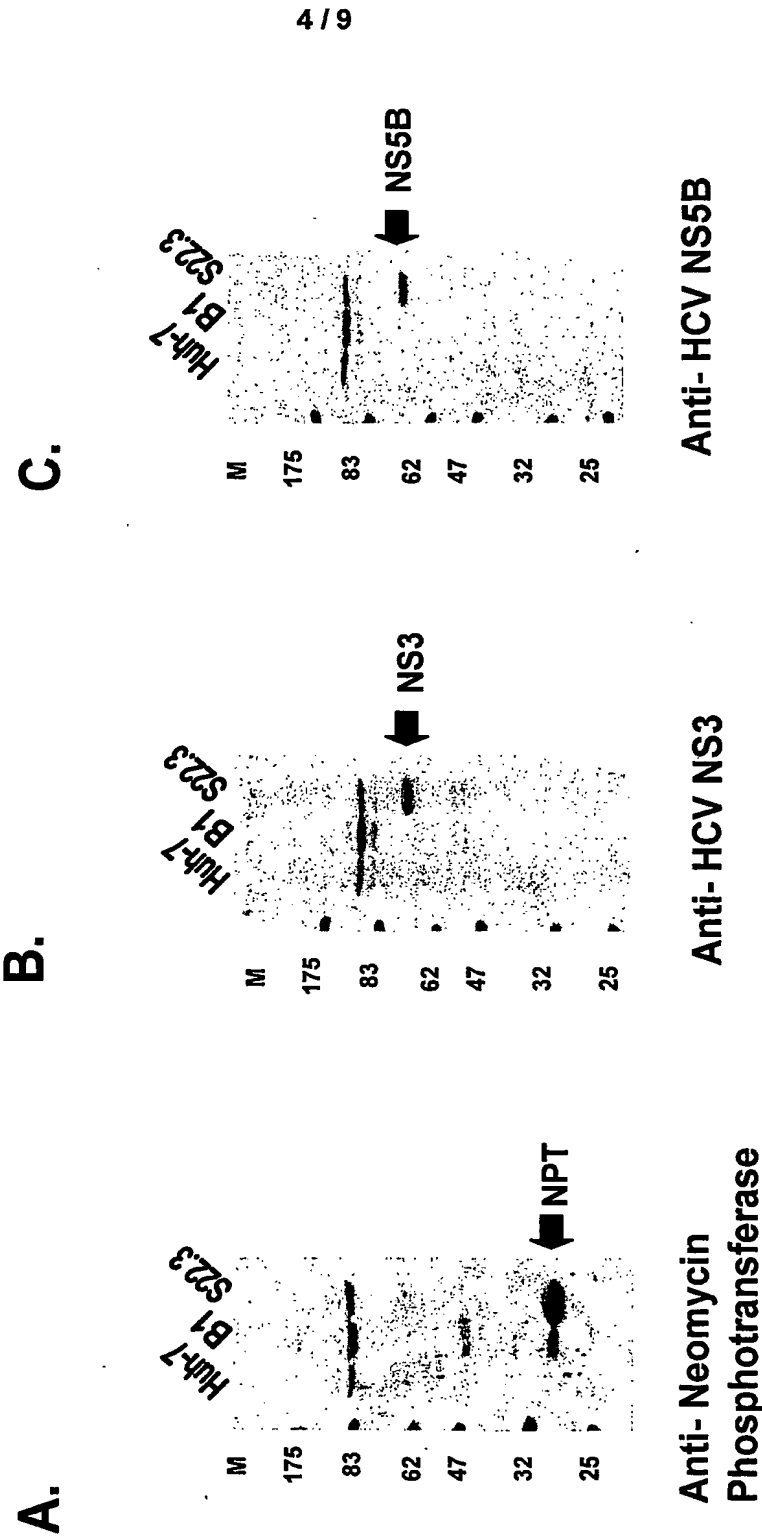


FIGURE 4



5 / 9

FIGURE 5A

	S 22-3 SEQ ID NO 2	R3 SEQ ID NO. 4	R3-rep SEQ ID NO. 7	R7 SEQ ID NO. 5	R16 SEQ ID NO 6
5'end - FIRST nt (HCV IRES)	*G (nt 1) A	G (nt 1) A	-	-	G (nt 1) A
Neo	-	A (nt 481) G	-	-	-
EMCV IRES	-	A (nt 1739) G	-	-	-
NS 2	-	-	-	-	-
NS 3	-	G (nt 2778) A A (nt 2840) C A (nt 4052) G	T (nt 2509) C G (nt 2778) A A (nt 2840) C T (nt 3574) C A (nt 4052) G	A (nt 2936) G A (nt 2979) G	A (nt 2816) G A (nt 2879) G
NS 4A	A (nt 4446) R	A (nt 4446) G	C (nt 4387) T A (nt 4446) G C (nt 4507) T	-	C (nt 4476) T
NS 4B	-	T (nt 4856) C	T (nt 4856) C	-	-
NS 5A	G (nt 5498) T A (nt 6268) R	A (nt 6351) G G (nt 5498) T G (nt 5659) A C (nt 5671) T A (nt 6268) G	A (nt 6351) G G (nt 5498) T G (nt 5659) A T (nt 5838) C C (nt 5871) T A (nt 6116) G	A (nt 6324) G G (nt 5498) T T (nt 6001) C	G (nt 5498) C T (nt 6320) C T (nt 6584) C
NS 5B	-	A (nt 6652) G	-	C (nt 7292) T T (nt 8349) C	-
3'end - last 98 nt	-	-	-	-	-

\*first nt = G from HCV ires

FIGURE 5B

	S 22-3 SEQ ID NO. 2	R3 SEQ ID NO. 4	R3 Rep SEQ ID NO. 7	R7 SEQ ID NO. 5	R16 SEQ ID NO. 6
5'end - FIRST nt (HCV IRES)	G (nt 1) A	G (nt 1) A	-	-	G (nt 1) A
NS 2	-	-	-	-	-
NS 3	-	R (1135) K S (1560) G	R (1135) K S (1560) G	E (1202) G	S (1148) G E (1202) G
NS 4A	K (1691) mix K/R	K (1691) R	K (1691) R	-	L (1701) F
NS 4B	-	-	-	-	-
NS 5A	G (2042) C	T (1993) A G (2042) C P (2166) L	T (1993) A G (2042) C L (2155) P P (2166) L	I (1984) V G (2042) C	G (2042) R S (2404) P
NS 5B	-	-	-	M (2992) T	-
3'end - last 98 nt	-	-	-	-	-

first a.a. of NS2 = 810

7 / 9

FIGURE 6

## AMINO ACID SUBSTITUTIONS

CLONE  
APGK-12

5' HCV IRES	NeoR	EMCV IRES	HCV NS2→5B				NS5B	3'HCV UTR
			NS2	NS3	4A	NS4B	NS5A	
G (nt1) SEQ ID NO 1								
A (nt1) SEQ ID NO 24								
R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1691)R	-	T(1993)A G(2042)C L(2155)P P(2166)L	
G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1691)R	-	T(1993)A G(2042)C L(2155)P P(2166)L	



77 cfu/μg



86 cfu/μg



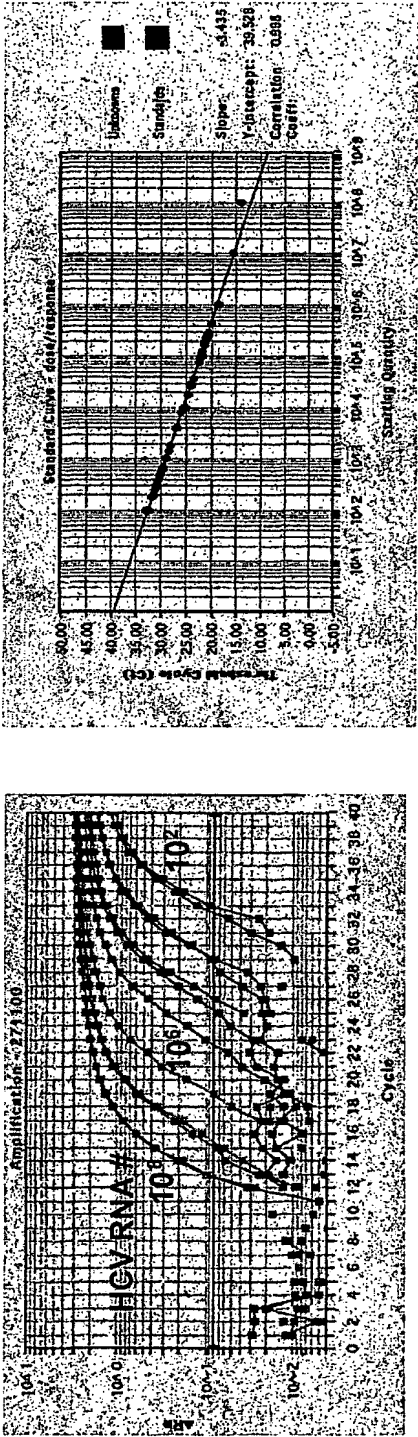
1100000cfu/μg



2000000cfu/μg

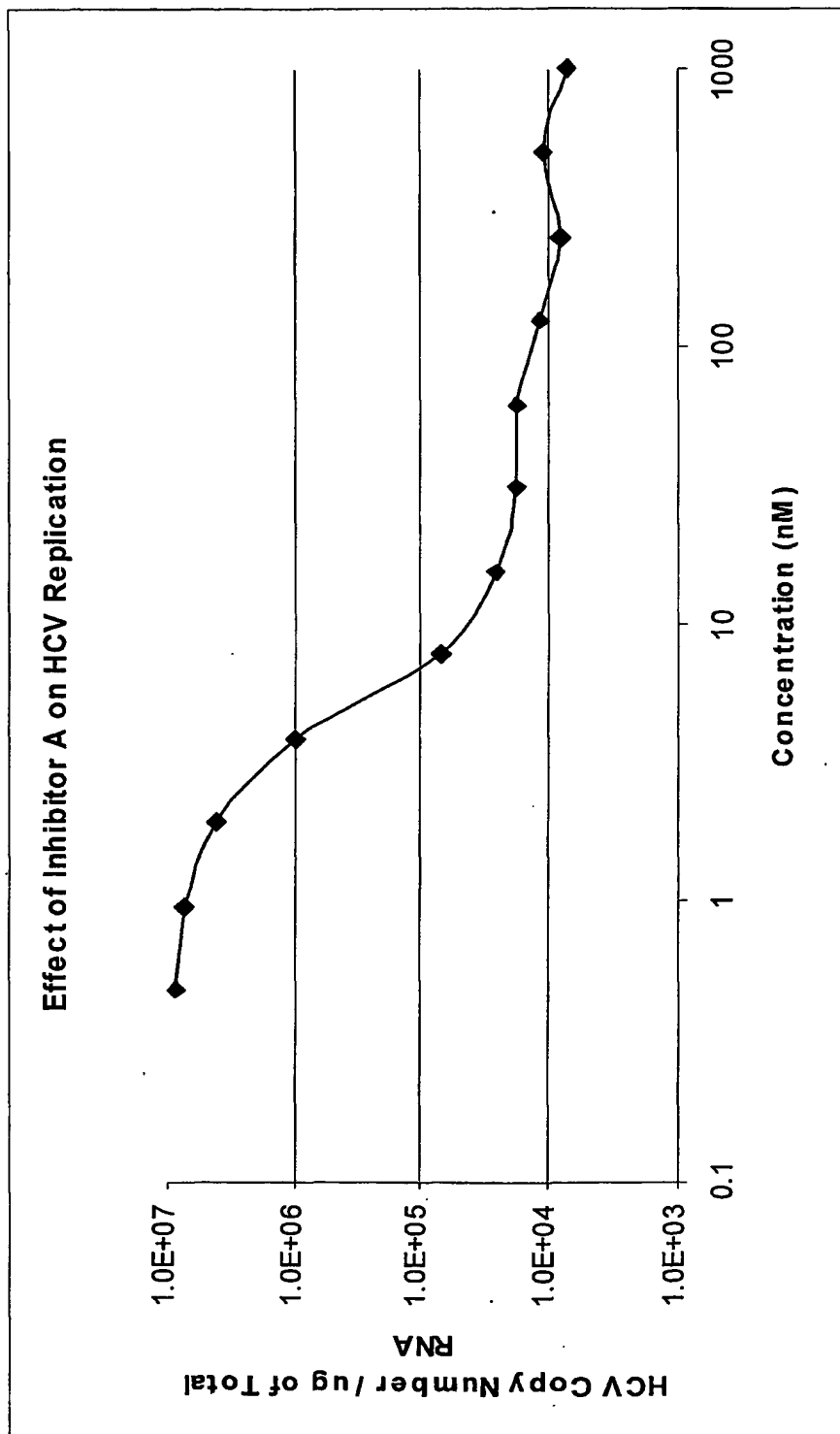
SUBSTITUTE SHEET (RULE 26)

FIGURE 7  
HCV-Replicon: RNA Quantification



Ct = Threshold cycle  $\propto$  Starting RNA Quantity

FIGURE 8





1 / 93

## SEQUENCE LISTING

&lt;110&gt; BOEHRINGER INGELHEIM (CANADA) LTD.

<120> SELF REPLICATING RNA MOLECULE FROM  
HEPATITIS C VIRUS

&lt;130&gt; 13/083

&lt;150&gt; 60/257,857

&lt;151&gt; 2000-12-22

&lt;160&gt; 25

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 8639

&lt;212&gt; DNA

&lt;213&gt; HCV

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1803) ... (8408)

&lt;400&gt; 1

```
ggccagcccc cgattggggg cgacactcca ccatagatca ctcccctgtg aggaactact 60
gtcttcacgc agaaagcgtc tagccatggc gttagtatga gtgtcgtgca gcctccagga 120
ccccccctcc cgggagagcc atagtggctc gcggaaccgg tgagtacacc ggaattgcca 180
ggacgaccgg gtcctttctt ggatcaacct gctcaatgcc tggagatttg ggcgtgcccc 240
cgcgagactg ctagccgagt agtggtgggt cgcgaaaggc cttgtgttac tgcctgatag 300
ggtgcttgcg agtgccccgg gaggtctcgt agaccgtgca ccatgagcac gaatcctaaa 360
cctcaaagaa aaaccaaagg gcgcgccatg attgaacaag atggattgca cgcaggttct 420
ccggccgctt ggggtggagag gctattcggc tatgactggg cacaacagac aatcggctgc 480
tctgatgcgc ccgtgttccg gctgtcagcg cagggggcgc cggttctttt tgtcaagacc 540
gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc 600
acgacgggcg ttccttgccg agctgtgtc gacgttgtca ctgaagcggg aagggactgg 660
ctgctatttg gcgaagtgcc ggggcaggat ctccgtgcat ctacacctgc tcctgccgag 720
aaagtatcca tcatggctga tgcaatgcgg cggtgcata cgcttgatcc ggctacctgc 780
ccattcgacc accaagcgaa acatcgcatc gagcagcac gtactcggat ggaagccggt 840
cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc 900
gccaggctca aggcgcgcac gcccgacggc gaggatctcg tcgtgaccca tggcgaatgc 960
tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg 1020
ctgggtgtgg cggaccgcta tcaggacata gcgttggtta cccgtgatat tgctgaagag 1080
cttggcggcg aatgggctga ccgcttctc gtgctttacg gtatcgccgc tcccgattcg 1140
cagcgcacgc ccttctatcg ccttcttgac gagttctctt gagttcgcgc ccagatgtta 1200
acagaccaca acggtttccc tctagcggga tcaattccgc cccccccct aacgttactg 1260
gccgaagccg cttggaataa ggccgggtgt cgtttgtcta tatgttattt tccaccatat 1320
tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg acgagcattc 1380
ctaggggtct tccccctctc gccaaaggaa tgcaaggctc gttgaatgtc gtgaaggaa 1440
cagttcctct ggaagcttct tgaagacaaa caacgtctgt agcgaccctt tgcaggcagc 1500
ggaaccccc cactggcgac aggtgcctct gcggccaaaa gccacgtgta taagatacac 1560
ctgcaaaggc ggcacaaccc cagtgccacg ttgtgagttg gatagttgtg gaaagagtca 1620
aatggctctc ctcaagcgta ttcaacaagg ggctgaagga tgcccagaag gtacccatt 1680
gtatgggatc tgatctgggg cctcgggtga catgctttac atgtgtttag tcgaggttaa 1740
aaaacgtcta ggcccccca accacgggga cgtggttttc ctttgaaaa cagcataata 1800
```

## 2 / 93

cc atg gac cgg gag atg gca gca tgg tgc gga ggc gcg gtt ttc gta	1847
Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val	
1 5 10 15	
ggc ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct	1895
Gly Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala	
20 25 30	
agg ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac	1943
Arg Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His	
35 40 45	
ttg caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc	1991
Leu Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Arg Asp Ala	
50 55 60	
gtc atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc	2039
Val Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile	
65 70 75	
acc aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct	2087
Thr Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala	
80 85 90 95	
ggc ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt	2135
Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg	
100 105 110	
gca tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg	2183
Ala Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met	
115 120 125	
gct ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat	2231
Ala Leu Met Lys Leu Ala Ala Thr Gly Thr Tyr Val Tyr Asp His	
130 135 140	
ctc acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg	2279
Leu Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala	
145 150 155	
gtg gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc	2327
Val Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile	
160 165 170 175	
acc tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg	2375
Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu	
180 185 190	
ccc gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac	2423
Pro Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp	
195 200 205	
agc ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac	2471
Ser Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr	
210 215 220	

## 3 / 93

tcc caa cag acg cga ggc cta ctt ggc tgc atc atc act agc ctc aca	2519
Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr	
225 230 235	
ggc cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc	2567
Gly Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr	
240 245 250 255	
gca aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act	2615
Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr	
260 265 270	
gtc tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca	2663
Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro	
275 280 285	
atc acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa	2711
Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln	
290 295 300	
gcg ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg	2759
Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser	
305 310 315	
gac ctt tac ttg gtc acg agg cat gcc gat gtc att ccg gtg cgc cgg	2807
Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg	
320 325 330 335	
cgg ggc gac agc agg ggg agc cta ctc tcc ccc agg ccc gtc tcc tac	2855
Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr	
340 345 350	
ttg aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct	2903
Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala	
355 360 365	
gtg ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg	2951
Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala	
370 375 380	
gtg gac ttt gta ccc gtc gag tct atg gaa acc act atg cgg tcc ccg	2999
Val Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro	
385 390 395	
gtc ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag	3047
Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln	
400 405 410 415	
gtg gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg	3095
Val Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val	
420 425 430	
ccg gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg	3143
Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro	
435 440 445	

## 4 / 93

tcc gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat	3191
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His	
450 455 460	
ggt atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt	3239
Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly	
465 470 475	
gcc ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt	3287
Ala Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly	
480 485 490 495	
tgc tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca	3335
Cys Ser Gly Gly Tyr Asp Ile Ile Cys Asp Glu Cys His Ser	
500 505 510	
act gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg	3383
Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala	
515 520 525	
gag acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg	3431
Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro	
530 535 540	
gga tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc	3479
Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser	
545 550 555	
agc act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc	3527
Ser Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr	
560 565 570 575	
atc aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt	3575
Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys	
580 585 590	
gat gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca	3623
Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala	
595 600 605	
tat tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc	3671
Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val	
610 615 620	
att gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc	3719
Ile Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe	
625 630 635	
gac tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc	3767
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe	
640 645 650 655	
agc ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac	3815
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp	
660 665 670	

## 5 / 93

gcg gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg	3863
Ala Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met	
675 680 685	
ggc att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc	3911
Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe	
690 695 700	
gat tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac	3959
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr	
705 710 715	
gag ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac	4007
Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn	
720 725 730 735	
aca cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag agc	4055
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser	
740 745 750	
gtc ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act	4103
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr	
755 760 765	
aag cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg	4151
Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr	
770 775 780	
gtg tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg	4199
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp	
785 790 795	
aag tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg	4247
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu	
800 805 810 815	
ctg tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc	4295
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro	
820 825 830	
ata acc aaa tac atc atg gca tgc atg tcg gct gac ctg gag gtc gtc	4343
Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val	
835 840 845	
acg agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg	4391
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala	
850 855 860	
tat tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg	4439
Tyr Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu	
865 870 875	
tcc gga aag ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag	4487
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu	
880 885 890 895	

6 / 93

ttc gat gag atg gaa gag tgc gcc tca cac ctc cct tac atc gaa cag	4535
Phe Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln	
900 905 910	
gga atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg	4583
Gly Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu	
915 920 925	
caa aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc	4631
Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala Pro Val Val Glu Ser	
930 935 940	
aag tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc	4679
Lys Trp Arg Thr Leu Glu Phe Trp Ala Lys His Met Trp Asn Phe	
945 950 955	
atc agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac	4727
Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn	
960 965 970 975	
ccc gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg	4775
Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro	
980 985 990	
ctc acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg	4823
Leu Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val	
995 1000 1005	
gcc gcc caa ctt gct cct ccc agc gct gct tct gct ttc gta ggc gcc	4871
Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala	
1010 1015 1020	
ggc atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt	4919
Gly Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu	
1025 1030 1035	
gtg gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg	4967
Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val	
1040 1045 1050 1055	
gcc ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt	5015
Ala Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val	
1060 1065 1070	
aac cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc	5063
Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val	
1075 1080 1085	
gtg tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct	5111
Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala	
1090 1095 1100	
gtg cag tgg atg aac cgg ctg ata gcg ttc gct tgg cgg ggt aac cac	5159
Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His	
1105 1110 1115	

7 / 93

gtc tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc	5207
Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val	
1120 1125 1130 1135	
act cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt	5255
Thr Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu	
1140 1145 1150	
cac cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg	5303
His Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp	
1155 1160 1165	
cta aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag	5351
Leu Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys	
1170 1175 1180	
acc tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc	5399
Thr Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe	
1185 1190 1195	
ttc tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc	5447
Phe Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile	
1200 1205 1210 1215	
atg caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa	5495
Met Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys	
1220 1225 1230	
aac ggt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg	5543
Asn Gly Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp	
1235 1240 1245	
cat gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc	5591
His Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro	
1250 1255 1260	
tcc ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag	5639
Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu	
1265 1270 1275	
gag tac gtg gag gtt acg cgg gtg ggg gat ttc cac tac gtg acg ggc	5687
Glu Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly	
1280 1285 1290 1295	
atg acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa	5735
Met Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu	
1300 1305 1310	
ttc ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg	5783
Phe Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala	
1315 1320 1325	
tgc aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat	5831
Cys Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn	
1330 1335 1340	

## 8 / 93

caa tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta	5879
Gln Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val	
1345 1350 1355	
gca gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag	5927
Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu	
1360 1365 1370 1375	
acg gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc	5975
Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser	
1380 1385 1390	
tca tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act	6023
Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr	
1395 1400 1405	
acc cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg	6071
Thr Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu	
1410 1415 1420	
tgg cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat	6119
Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn	
1425 1430 1435	
aag gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat	6167
Lys Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp	
1440 1445 1450 1455	
gag agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa	6215
Glu Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys	
1460 1465 1470	
ttc cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca	6263
Phe Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro	
1475 1480 1485	
ctg tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac	6311
Leu Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His	
1490 1495 1500	
ggg tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg	6359
Gly Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg	
1505 1510 1515	
agg aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg	6407
Arg Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu	
1520 1525 1530 1535	
gcg gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc	6455
Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val	
1540 1545 1550	
gac agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc	6503
Asp Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly	
1555 1560 1565	



## 9 / 93

gac gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag	6551
Asp Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu	
1570 1575 1580	
ggg gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta	6599
Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val	
1585 1590 1595	
agc gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca	6647
Ser Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr	
1600 1605 1610 1615	
tgg aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg	6695
Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu	
1620 1625 1630	
ccc atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc	6743
Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val	
1635 1640 1645	
tat gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc	6791
Tyr Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr	
1650 1655 1660	
ttt gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag	6839
Phe Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys	
1665 1670 1675	
gag atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg	6887
Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val	
1680 1685 1690 1695	
gag gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt	6935
Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe	
1700 1705 1710	
ggc tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac	6983
Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn	
1715 1720 1725	
cac atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca	7031
His Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro	
1730 1735 1740	
att gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca	7079
Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro	
1745 1750 1755	
gag aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg	7127
Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu	
1760 1765 1770 1775	
ggg gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc	7175
Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr	
1780 1785 1790	

.10 / 93

ctc cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct	7223
Leu Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro	
1795 1800 1805	
gga cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc	7271
Gly Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys	
1810 1815 1820	
cct atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act	7319
Pro Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr	
1825 1830 1835	
gag aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg	7367
Glu Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu	
1840 1845 1850 1855	
gcc ccc gaa gcc aga cag gcc ata agg tgc ctc aca gag cgg ctt tac	7415
Ala Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr	
1860 1865 1870	
atc ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc	7463
Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg	
1875 1880 1885	
cgg tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc	7511
Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu	
1890 1895 1900	
aca tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag	7559
Thr Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln	
1905 1910 1915	
gac tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa	7607
Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu	
1920 1925 1930 1935	
agc gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag	7655
Ser Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu	
1940 1945 1950	
gct atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa	7703
Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu	
1955 1960 1965	
tac gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg	7751
Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala	
1970 1975 1980	
cac gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc	7799
His Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr	
1985 1990 1995	
acc ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc	7847
Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val	
2000 2005 2010 2015	

## 11 / 93

```

aat tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca 7895
Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala
      2020      2025      2030

agg atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa 7943
Arg Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu
      2035      2040      2045

caa ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc 7991
Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser
      2050      2055      2060

att gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt 8039
Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu
      2065      2070      2075

agc gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg 8087
Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val
      2080      2085      2090      2095

gct tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga 8135
Ala Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg
      2100      2105      2110

cat cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg 8183
His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg
      2115      2120      2125

gct gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag 8231
Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys
      2130      2135      2140

ctc aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc 8279
Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser
      2145      2150      2155

tgg ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct 8327
Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser
      2160      2165      2170      2175

cgt gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta 8375
Arg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val
      2180      2185      2190

ggg gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8428
Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg *
      2195      2200

ggccaatagg ccatacctgtt tttttccctt tttttttttt tttttttttt tttttttttt 8488
tttttttttt tttttctcctt tttttttcct cttttttttcc tttttctttcc ttttggtggct 8548
ccatccttagc cctagtcacg gctagctgtg aaagggtccgt gagccgcttg actgcagaga 8608
gtgctgatac tggcctctct gcagatcaag t 8639

```

&lt;210&gt; 2

&lt;211&gt; 8642

&lt;212&gt; DNA

&lt;213&gt; HCV

&lt;220&gt;

12 / 93

<221> CDS  
 <222> (1802)...(8407)

<221> variation  
 <222> 6268  
 <223> r = a or g  
 <221> variation  
 <222> 4446  
 <223> r = a or g

<400> 2  
 accagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60  
 tctttcacgca gaaagcgtct agccatggcg ttagtatgag tgctcgtgcag cctccaggac 120  
 ccccccctccc gggagagcca tagtggtctg cggaaccggg gactacaccg gaattgccag 180  
 gacgaccggg tcctttcttg gatcaaccgg ctcaatgcct ggagatttgg gcgtgcccc 240  
 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgggtact gcctgatagg 300  
 gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360  
 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420  
 cggccgccttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480  
 ctgatgcgcg cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540  
 acctgtccgg tgcctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600  
 cgacgggctg tccttgccga gctgtgctcg acgttgtcac tgaagcggga agggactggc 660  
 tgctatttgg cgaagtgcgg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720  
 aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780  
 cattogacca ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc 840  
 ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900  
 ccaggctcaa ggcgcgcag cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960  
 gcttgccgaa tatcatggtg gaaaatggcc gcttttcttg attcatcgac tgtggccggc 1020  
 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080  
 ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140  
 agcgcacatgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200  
 cagaccacaa cggtttccct ctacgaggat caattccgcc ccccccccta acgttactgg 1260  
 ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320  
 gccgtctttt ggcaatgtga gggcccgga acctggccct gtcttcttga cgagcattcc 1380  
 taggggtctt tccctctcg ccaaaggaat gcaaggctcg ttgaatgtcg tgaaggagc 1440  
 agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgaccttt gcaggcagcg 1500  
 gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560  
 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620  
 atggctctcc tcaagcgtat tcaacaagg gctgaaggat gccagaagg taccctattg 1680  
 tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740  
 aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800  
 c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849  
 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly  
 1 5 10 15  
 ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897  
 Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg  
 20 25 30  
 ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 1945  
 Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu  
 35 40 45  
 caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993  
 Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val  
 50 55 60

## 13 / 93

atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc	2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr	
65 70 75 80	
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt	2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly	
85 90 95	
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca	2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala	
100 105 110	
tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct	2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala	
115 120 125	
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc	2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu	
130 135 140	
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg	2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val	
145 150 155 160	
gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc	2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr	
165 170 175	
tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc	2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro	
180 185 190	
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc	2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser	
195 200 205	
ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc act agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	
cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	

## 14 / 93

acc	caa	atg	tac	acc	aat	gtg	gac	cag	gac	ctc	gtc	ggc	tgg	caa	gcg	2713
Thr	Gln	Met	Tyr	Thr	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	Gln	Ala	
290						295					300					
ccc	ccc	ggg	gcg	cgt	tcc	ttg	aca	cca	tgc	acc	tgc	ggc	agc	tcg	gac	2761
Pro	Pro	Gly	Ala	Arg	Ser	Leu	Thr	Pro	Cys	Thr	Cys	Gly	Ser	Ser	Asp	
305					310					315					320	
ctt	tac	ttg	gtc	acg	agg	cat	gcc	gat	gtc	att	ccg	gtg	cgc	cgg	cgg	2809
Leu	Tyr	Leu	Val	Thr	Arg	His	Ala	Asp	Val	Ile	Pro	Val	Arg	Arg	Arg	
				325					330					335		
ggc	gac	agc	agg	ggg	agc	cta	ctc	tcc	ccc	agg	ccc	gtc	tcc	tac	ttg	2857
Gly	Asp	Ser	Arg	Gly	Ser	Leu	Leu	Ser	Pro	Arg	Pro	Val	Ser	Tyr	Leu	
			340					345					350			
aag	ggc	tct	tcg	ggc	ggc	cca	ctg	ctc	tgc	ccc	tcg	ggg	cac	gct	gtg	2905
Lys	Gly	Ser	Ser	Gly	Gly	Pro	Leu	Leu	Cys	Pro	Ser	Gly	His	Ala	Val	
		355					360					365				
ggc	atc	ttt	cgg	gct	gcc	gtg	tgc	acc	cga	ggg	ggt	gcg	aag	gcg	gtg	2953
Gly	Ile	Phe	Arg	Ala	Ala	Val	Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	
	370					375					380					
gac	ttt	gta	ccc	gtc	gag	tct	atg	gaa	acc	act	atg	cgg	tcc	ccg	gtc	3001
Asp	Phe	Val	Pro	Val	Glu	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	
385					390					395					400	
ttc	acg	gac	aac	tcg	tcc	cct	ccg	gcc	gta	ccg	cag	aca	ttc	cag	gtg	3049
Phe	Thr	Asp	Asn	Ser	Ser	Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val	
			405						410					415		
gcc	cat	cta	cac	gcc	cct	act	ggc	agc	ggc	aag	agc	act	aag	gtg	ccg	3097
Ala	His	Leu	His	Ala	Pro	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Lys	Val	Pro	
			420				425						430			
gct	gcg	tat	gca	gcc	caa	ggg	tat	aag	gtg	ctt	gtc	ctg	aac	ccg	tcc	3145
Ala	Ala	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	
		435				440						445				
gtc	gcc	gcc	acc	cta	ggc	ttc	ggg	gcg	tat	atg	tct	aag	gca	cat	ggc	3193
Val	Ala	Ala	Thr	Leu	Gly	Phe	Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	
	450				455					460						
atc	gac	cct	aac	atc	aga	acc	ggg	gta	agg	acc	atc	acc	acg	ggc	gcc	3241
Ile	Asp	Pro	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala	
465					470					475					480	
ccc	atc	acg	tac	tcc	acc	tat	ggc	aag	ttt	ctt	gcc	gac	ggc	ggc	tgc	3289
Pro	Ile	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	
				485					490					495		
tct	ggg	ggc	gcc	tat	gac	atc	ata	ata	tgt	gat	gag	tgc	cac	tca	act	3337
Ser	Gly	Gly	Ala	Tyr	Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	
			500					505					510			

## 15 / 93

gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	
act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	
ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	
att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	

## 16 / 93

cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag agc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tgc tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tgc gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga arg ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc	4489
Gly Xaa Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgc gcc tca cac ctg cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	
atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	



## 17 / 93

```

agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc 4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro
          965                      970                      975

gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc 4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu
          980                      985                      990

acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc 4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala
          995                      1000                      1005

gcc caa ctt gct cct ccc agc gct gct tct gct ttc gta ggc gcc ggc 4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly
          1010                      1015                      1020

atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg 4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val
          1025                      1030                      1035                      1040

gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc 4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala
          1045                      1050                      1055

ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac 5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn
          1060                      1065                      1070

cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg 5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val
          1075                      1080                      1085

tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg 5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val
          1090                      1095                      1100

cag tgg atg aac cgg ctg ata gcg ttc gct tgc cgg ggt aac cac gtc 5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val
          1105                      1110                      1115                      1120

tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act 5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr
          1125                      1130                      1135

cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac 5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His
          1140                      1145                      1150

cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg cta 5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu
          1155                      1160                      1165

aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag acc 5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Thr
          1170                      1175                      1180

```

## 18 / 93

tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc ttg ccg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
tgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg ttg ccg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg ccg gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg ccg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta ccg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta gca	5881
Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala	
1345 1350 1355 1360	
gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	

## 19 / 93

cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
ttr gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Xaa Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	
gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	

20 / 93

gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	

## 21 / 93

ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn	
2005 2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg	
2020 2025 2030	
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa	7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln	
2035 2040 2045	
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att	7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile	
2050 2055 2060	
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc	8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser	
2065 2070 2075 2080	
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct	8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala	
2085 2090 2095	

22 / 93

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137  
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His  
2100 2105 2110

cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185  
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala  
2115 2120 2125

gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233  
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu  
2130 2135 2140

aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281  
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp  
2145 2150 2155 2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329  
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg  
2165 2170 2175

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 8377  
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly  
2180 2185 2190

gta ggc atc tat cta ctc ccc aac cga tga acgaggagct aaacactcca 8427  
Val Gly Ile Tyr Leu Leu Pro Asn Arg \*  
2195 2200

ggccaatagg	ccatcctgtt	tttttccoct	tttttttttt	tttttttttc	tttttttttt	8487
tttttttttt	tttttttttc	tccttttttt	tcctcttttt	ttccttttct	ttccttgggt	8547
ggctccatct	tagccctagt	cacggctagc	tgtgaaaggt	ccgtgagccg	cttgactgca	8607
gagagtgctg	atactggcct	ctctgcagat	caagt			8642

<210> 3

<211> 2201

<212> PRT

<213> HCV

<220>

<221> VARIANT

<222> 882

<223> Xaa is Lys or Arg

<221> VARIANT

<222> 1489

<223> Xaa is Leu

<400> 3

Met	Asp	Arg	Glu	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly
1				5					10					15	
Leu	Ile	Leu	Leu	Thr	Leu	Ser	Pro	His	Tyr	Lys	Leu	Phe	Leu	Ala	Arg
			20					25					30		
Leu	Ile	Trp	Trp	Leu	Gln	Tyr	Phe	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu
		35					40					45			
Gln	Val	Trp	Ile	Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Val
		50				55					60				
Ile	Leu	Leu	Thr	Cys	Ala	Ile	His	Pro	Glu	Leu	Ile	Phe	Thr	Ile	Thr
65					70					75					80

23 / 93

Lys	Ile	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Gly		
				85					90					95			
Ile	Thr	Lys	Val	Pro	Tyr	Phe	Val	Arg	Ala	His	Gly	Leu	Ile	Arg	Ala		
			100					105					110				
Cys	Met	Leu	Val	Arg	Lys	Val	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala		
			115					120					125				
Leu	Met	Lys	Leu	Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Asp	His	Leu		
			130					135				140					
Thr	Pro	Leu	Arg	Asp	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val		
					150					155					160		
Ala	Val	Glu	Pro	Val	Val	Phe	Ser	Asp	Met	Glu	Thr	Lys	Val	Ile	Thr		
				165					170						175		
Trp	Gly	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro		
			180					185					190				
Val	Ser	Ala	Arg	Arg	Gly	Arg	Glu	Ile	His	Leu	Gly	Pro	Ala	Asp	Ser		
			195				200					205					
Leu	Glu	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ser		
			210			215					220						
Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly		
				230						235					240		
Arg	Asp	Arg	Asn	Gln	Val	Glu	Gly	Glu	Val	Gln	Val	Val	Ser	Thr	Ala		
			245					250							255		
Thr	Gln	Ser	Phe	Leu	Ala	Thr	Cys	Val	Asn	Gly	Val	Cys	Trp	Thr	Val		
			260					265					270				
Tyr	His	Gly	Ala	Gly	Ser	Lys	Thr	Leu	Ala	Gly	Pro	Lys	Gly	Pro	Ile		
			275				280					285					
Thr	Gln	Met	Tyr	Thr	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp	Gln	Ala		
			290			295					300						
Pro	Pro	Gly	Ala	Arg	Ser	Leu	Thr	Pro	Cys	Thr	Cys	Gly	Ser	Ser	Asp		
					310					315					320		
Leu	Tyr	Leu	Val	Thr	Arg	His	Ala	Asp	Val	Ile	Pro	Val	Arg	Arg	Arg		
				325					330						335		
Gly	Asp	Ser	Arg	Gly	Ser	Leu	Leu	Ser	Pro	Arg	Pro	Val	Ser	Tyr	Leu		
			340					345					350				
Lys	Gly	Ser	Ser	Gly	Gly	Pro	Leu	Leu	Cys	Pro	Ser	Gly	His	Ala	Val		
			355				360					365					
Gly	Ile	Phe	Arg	Ala	Ala	Val	Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val		
			370			375					380						
Asp	Phe	Val	Pro	Val	Glu	Ser	Met	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val		
				390					395						400		
Phe	Thr	Asp	Asn	Ser	Ser	Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val		
			405						410						415		
Ala	His	Leu	His	Ala	Pro	Thr	Gly	Ser	Gly	Lys	Ser	Thr	Lys	Val	Pro		
			420					425					430				
Ala	Ala	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser		
			435				440					445					
Val	Ala	Ala	Thr	Leu	Gly	Phe	Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly		
			450			455					460						
Ile	Asp	Pro	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ala		
				470						475					480		
Pro	Ile	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys		
			485						490						495		
Ser	Gly	Gly	Ala	Tyr	Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr		
			500					505					510				
Asp	Ser	Thr	Ile	Leu	Gly	Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu			
			515			520						525					
Thr	Ala	Gly	Ala	Arg	Leu	Val	Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly		
			530			535						540					

## 24 / 93

Ser	Val	Thr	Val	Pro	His	Pro	Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Ser	545	550	555	560
Thr	Gly	Glu	Ile	Pro	Phe	Tyr	Gly	Lys	Ala	Ile	Pro	Ile	Glu	Thr	Ile	565	570	575	
Lys	Gly	Gly	Arg	His	Leu	Ile	Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	580	585	590	
Glu	Leu	Ala	Ala	Lys	Leu	Ser	Gly	Leu	Gly	Leu	Asn	Ala	Val	Ala	Tyr	595	600	605	
Tyr	Arg	Gly	Leu	Asp	Val	Ser	Val	Ile	Pro	Thr	Ser	Gly	Asp	Val	Ile	610	615	620	
Val	Val	Ala	Thr	Asp	Ala	Leu	Met	Thr	Gly	Phe	Thr	Gly	Asp	Phe	Asp	625	630	635	640
Ser	Val	Ile	Asp	Cys	Asn	Thr	Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	645	650	655	
Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr	Thr	Thr	Val	Pro	Gln	Asp	Ala	660	665	670	
Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Met	Gly	675	680	685	
Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe	Asp	690	695	700	
Ser	Ser	Val	Leu	Cys	Glu	Cys	Tyr	Asp	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	705	710	715	720
Leu	Thr	Pro	Ala	Glu	Thr	Ser	Val	Arg	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	725	730	735	
Pro	Gly	Leu	Pro	Val	Cys	Gln	Asp	His	Leu	Glu	Phe	Trp	Glu	Ser	Val	740	745	750	
Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	His	Phe	Leu	Ser	Gln	Thr	Lys	755	760	765	
Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	770	775	780	
Cys	Ala	Arg	Ala	Gln	Ala	Pro	Pro	Pro	Ser	Trp	Asp	Gln	Met	Trp	Lys	785	790	795	800
Cys	Leu	Ile	Arg	Leu	Lys	Pro	Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	805	810	815	
Tyr	Arg	Leu	Gly	Ala	Val	Gln	Asn	Glu	Val	Thr	Thr	Thr	His	Pro	Ile	820	825	830	
Thr	Lys	Tyr	Ile	Met	Ala	Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	835	840	845	
Ser	Thr	Trp	Val	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	850	855	860	
Cys	Leu	Thr	Thr	Gly	Ser	Val	Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	865	870	875	880
Gly	Xaa	Pro	Ala	Ile	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Arg	Glu	Phe	885	890	895	
Asp	Glu	Met	Glu	Glu	Cys	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	900	905	910	
Met	Gln	Leu	Ala	Glu	Gln	Phe	Lys	Gln	Lys	Ala	Ile	Gly	Leu	Leu	Gln	915	920	925	
Thr	Ala	Thr	Lys	Gln	Ala	Glu	Ala	Ala	Ala	Pro	Val	Val	Glu	Ser	Lys	930	935	940	
Trp	Arg	Thr	Leu	Glu	Ala	Phe	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	945	950	955	960
Ser	Gly	Ile	Gln	Tyr	Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro	Gly	Asn	Pro	965	970	975	
Ala	Ile	Ala	Ser	Leu	Met	Ala	Phe	Thr	Ala	Ser	Ile	Thr	Ser	Pro	Leu	980	985	990	
Thr	Thr	Gln	His	Thr	Leu	Leu	Phe	Asn	Ile	Leu	Gly	Gly	Trp	Val	Ala	995	1000	1005	



25 / 93

Ala	Gln	Leu	Ala	Pro	Pro	Ser	Ala	Ala	Ser	Ala	Phe	Val	Gly	Ala	Gly	1010	1015	1020
Ile	Ala	Gly	Ala	Ala	Val	Gly	Ser	Ile	Gly	Leu	Gly	Lys	Val	Leu	Val	1025	1030	1035
Asp	Ile	Leu	Ala	Gly	Tyr	Gly	Ala	Gly	Val	Ala	Gly	Ala	Leu	Val	Ala	1045	1050	1055
Phe	Lys	Val	Met	Ser	Gly	Glu	Met	Pro	Ser	Thr	Glu	Asp	Leu	Val	Asn	1060	1065	1070
Leu	Leu	Pro	Ala	Ile	Leu	Ser	Pro	Gly	Ala	Leu	Val	Val	Gly	Val	Val	1075	1080	1085
Cys	Ala	Ala	Ile	Leu	Arg	Arg	His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	1090	1095	1100
Gln	Trp	Met	Asn	Arg	Leu	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val	1105	1110	1115
Ser	Pro	Thr	His	Tyr	Val	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr	1125	1130	1135
Gln	Ile	Leu	Ser	Ser	Leu	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His	1140	1145	1150
Gln	Trp	Ile	Asn	Glu	Asp	Cys	Ser	Thr	Pro	Cys	Ser	Gly	Ser	Trp	Leu	1155	1160	1165
Arg	Asp	Val	Trp	Asp	Trp	Ile	Cys	Thr	Val	Leu	Thr	Asp	Phe	Lys	Thr	1170	1175	1180
Trp	Leu	Gln	Ser	Lys	Leu	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe	1185	1190	1195
Ser	Cys	Gln	Arg	Gly	Tyr	Lys	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	1205	1210	1215
Gln	Thr	Thr	Cys	Pro	Cys	Gly	Ala	Gln	Ile	Thr	Gly	His	Val	Lys	Asn	1220	1225	1230
Cys	Ser	Met	Arg	Ile	Val	Gly	Pro	Arg	Thr	Cys	Ser	Asn	Thr	Trp	His	1235	1240	1245
Gly	Thr	Phe	Pro	Ile	Asn	Ala	Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Ser	1250	1255	1260
Pro	Ala	Pro	Asn	Tyr	Ser	Arg	Ala	Leu	Trp	Arg	Val	Ala	Ala	Glu	Glu	1265	1270	1275
Tyr	Val	Glu	Val	Thr	Arg	Val	Gly	Asp	Phe	His	Tyr	Val	Thr	Gly	Met	1285	1290	1295
Thr	Thr	Asp	Asn	Val	Lys	Cys	Pro	Cys	Gln	Val	Pro	Ala	Pro	Glu	Phe	1300	1305	1310
Phe	Thr	Glu	Val	Asp	Gly	Val	Arg	Leu	His	Arg	Tyr	Ala	Pro	Ala	Cys	1315	1320	1325
Lys	Pro	Leu	Leu	Arg	Glu	Glu	Val	Thr	Phe	Leu	Val	Gly	Leu	Asn	Gln	1330	1335	1340
Tyr	Leu	Val	Gly	Ser	Gln	Leu	Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Ala	1345	1350	1355
Val	Leu	Thr	Ser	Met	Leu	Thr	Asp	Pro	Ser	His	Ile	Thr	Ala	Glu	Thr	1365	1370	1375
Ala	Lys	Arg	Arg	Leu	Ala	Arg	Gly	Ser	Pro	Pro	Ser	Leu	Ala	Ser	Ser	1380	1385	1390
Ser	Ala	Ser	Gln	Leu	Ser	Ala	Pro	Ser	Leu	Lys	Ala	Thr	Cys	Thr	Thr	1395	1400	1405
Arg	His	Asp	Ser	Pro	Asp	Ala	Asp	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	1410	1415	1420
Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	1425	1430	1435
Val	Val	Ile	Leu	Asp	Ser	Phe	Glu	Pro	Leu	Gln	Ala	Glu	Glu	Asp	Glu	1445	1450	1455
Arg	Glu	Val	Ser	Val	Pro	Ala	Glu	Ile	Leu	Arg	Arg	Ser	Arg	Lys	Phe	1460	1465	1470

## 26 / 93

Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu  
 1475 1480 1485  
 Xaa Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly  
 1490 1495 1500  
 Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg  
 1505 1510 1515 1520  
 Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala  
 1525 1530 1535  
 Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp  
 1540 1545 1550  
 Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp  
 1555 1560 1565  
 Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly  
 1570 1575 1580  
 Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser  
 1585 1590 1595 1600  
 Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp  
 1605 1610 1615  
 Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro  
 1620 1625 1630  
 Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr  
 1635 1640 1645  
 Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe  
 1650 1655 1660  
 Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu  
 1665 1670 1675 1680  
 Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu  
 1685 1690 1695  
 Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly  
 1700 1705 1710  
 Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His  
 1715 1720 1725  
 Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile  
 1730 1735 1740  
 Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu  
 1745 1750 1755 1760  
 Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly  
 1765 1770 1775  
 Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu  
 1780 1785 1790  
 Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly  
 1795 1800 1805  
 Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro  
 1810 1815 1820  
 Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu  
 1825 1830 1835 1840  
 Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala  
 1845 1850 1855  
 Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile  
 1860 1865 1870  
 Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg  
 1875 1880 1885  
 Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr  
 1890 1895 1900  
 Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp  
 1905 1910 1915 1920  
 Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser  
 1925 1930 1935

27 / 93

Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala  
 1940 1945 1950  
 Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr  
 1955 1960 1965  
 Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His  
 1970 1975 1980  
 Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr  
 1985 1990 1995 2000  
 Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn  
 2005 2010 2015  
 Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg  
 2020 2025 2030  
 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln  
 2035 2040 2045  
 Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile  
 2050 2055 2060  
 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser  
 2065 2070 2075 2080  
 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala  
 2085 2090 2095  
 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His  
 2100 2105 2110  
 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala  
 2115 2120 2125  
 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu  
 2130 2135 2140  
 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp  
 2145 2150 2155 2160  
 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg  
 2165 2170 2175  
 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly  
 2180 2185 2190  
 Val Gly Ile Tyr Leu Leu Pro Asn Arg  
 2195 2200

&lt;210&gt; 4

&lt;211&gt; 8643

&lt;212&gt; DNA

&lt;213&gt; HCV

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1802)... (8407)

&lt;400&gt; 4

```

accagccccc gattgggggc gacaatccac catagatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgctgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gactacaccg gaattgccag 180
gacgaccggg tcctttcttg gatcaaccgc ctcaatgcct ggagatttgg gcgtgcccc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc gcaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780

```

28 / 93

```

cattcgacca ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatgggt gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tcttttacgg tatcgccgt cccgattcgc 1140
agcgcacatcg cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200
cagaccacaa cggtttccct ctacgaggat caattccgcc ccccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccgga acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggctcg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaagg gctgaaggat gcccagaagg taccatttg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaga 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
1 5 10 15

ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
20 25 30

ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
35 40 45

caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
50 55 60

atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc 2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
65 70 75 80

aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt 2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
85 90 95

ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca 2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
100 105 110

tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct 2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala
115 120 125

ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc 2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu
130 135 140

acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg 2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val
145 150 155 160

```

## 29 / 93

gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc	2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr	
165 170 175	
tggtgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc	2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro	
180 185 190	
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc	2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser	
195 200 205	
ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc act agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	
cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg aag cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Lys His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	
ggc gac agc agg ggg agc cta ctc tcc ccc cgg ccc gtc tcc tac ttg	2857
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	

## 30 / 93

gac ttt gta ccc gtc gag tct atg gaa acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	
gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	
atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc	3241
Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala	
465 470 475 480	
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc	3289
Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys	
485 490 495	
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act	3337
Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr	
500 505 510	
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	
act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	

## 31 / 93

tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	
ctg gac ccg acc ttc acc att gag acg acc acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	
att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag ggc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	

## 32 / 93

acc aaa tac atc atg gca tgc atg tgc gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga agg ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc	4489
Gly Arg Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgc gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	
atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	
gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	
atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	



## 33 / 93

ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	
cag tgg atg aac cgg ctg ata gcg ttc gct tgc cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	
cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag gcc	5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Ala	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
tgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	

## 34 / 93

tac gtg gag gtt acg cga gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ctg gac gta gca	5881
Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Leu Asp Val Ala	
1345 1350 1355 1360	
gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
ttg gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	

## 35 / 93

tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	
gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
acg ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	

## 36 / 93

atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	

## 37 / 93

atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849
Pro Leu Ala Arg Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn	
2005 2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg	
2020 2025 2030	
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa	7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln	
2035 2040 2045	
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att	7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile	
2050 2055 2060	
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc	8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser	
2065 2070 2075 2080	
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct	8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala	
2085 2090 2095	
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat	8137
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His	
2100 2105 2110	
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct	8185
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala	
2115 2120 2125	
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc	8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu	
2130 2135 2140	
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg	8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp	
2145 2150 2155 2160	
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt	8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg	
2165 2170 2175	

## 38 / 93

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 8377  
 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly  
 2180 2185 2190

gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427  
 Val Gly Ile Tyr Leu Leu Pro Asn Arg \*  
 2195 2200

ggccaatagg ccatoctggt ttttcccttt tttttttttt tttttttttt tttttttttt 8487  
 tttttttttt tttttttttt ttttcttttt tcccaatttt tttccctttt tttccctttg 8547  
 tggctccatc ttagccctag tcacggctag ctgtgaaagg tccgtgagcc gcttgactgc 8607  
 agagagtgtc gatactggcc tctctgcaga tcaagt 8643

<210> 5

<211> 8648

<212> DNA

<213> HCV

<220>

<221> CDS

<222> (1802)... (8407)

<400> 5

gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60  
 tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgctgtgcag cctccaggac 120  
 cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180  
 gacgaccggg tcctttcttg gatcaaccgc ctcaatgcct ggagatttgg gctgcccccc 240  
 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgggtact gctgatagg 300  
 gtgcttgca gtgccccggg aggtctcgta gaccgtgcac catgagcac aatcctaaac 360  
 ctcaaagaaa aaccaaagg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420  
 cgcccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggtgtgt 480  
 ctgatgccgc cgtgttccgg ctgtcagcgc agggggcgcc gggtcttttt gtcaagaccg 540  
 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600  
 cgacgggctg tccttgcgca gctgtgctcg acgttgctac tgaagcggga agggactggc 660  
 tgctattggg cgaagtgcgc gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720  
 aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatcgc gctacctgcc 780  
 cattcgacca ccaagcgaac catcgcatcg agcgagcacg tactcggatg gaagccgggtc 840  
 ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttgc 900  
 ccaggctcaa ggcgcgcctg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960  
 gcttgccgaa tatcatggtg gaaaatggcc gcttttcttg attcatcgac tgtggccggc 1020  
 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080  
 ttggcggcga atgggctgac cgcttctcgc tgctttacgg tatcgccgct cccgattcgc 1140  
 agcgcatcgc cttctatcgc cttcttgacg agttctctcg agttcgcgc cagatgttaa 1200  
 cagaccacaa cggtttccct ctacgggat caattccgcc ccccccccta acgttactgg 1260  
 ccgaagccgc ttggaataag gccggtgtgc gttgtctat atgttatttt ccaccatatt 1320  
 gccgtctttt ggcaatgtga gggcccgga acctggccct gtcttcttga cgaagcattcc 1380  
 taggggtctt tcccctctcg ccaaaggaat gcaaggctcg ttgaatgtcg tgaaggaagc 1440  
 agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgaccttt gcaggcagcg 1500  
 gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560  
 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620  
 atggctctcc tcaagcgtat tcaacaagg gctgaaggat gccagaagg taccocattg 1680  
 tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740  
 aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800  
 c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849  
 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly  
 1 5 10 15

## 39 / 93

ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg	1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg	
20 25 30	
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg	1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu	
35 40 45	
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg gcc cgc gat gcc gtc	1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val	
50 55 60	
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc	2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr	
65 70 75 80	
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt	2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly	
85 90 95	
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca	2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala	
100 105 110	
tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct	2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala	
115 120 125	
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc	2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu	
130 135 140	
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg	2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val	
145 150 155 160	
gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc	2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr	
165 170 175	
tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc	2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro	
180 185 190	
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc	2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser	
195 200 205	
ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc act agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	

## 40 / 93

cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg agg cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	
ggc gac agc agg ggg agc cta ctc tcc ccc agg ccc gtc tcc tac ttg	2857
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cgg ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	
gac ttt gta ccc gtc gag tct atg gga acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Gly Thr Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	
gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	



## 41 / 93

atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala 465 470 475 480	3241
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys 485 490 495	3289
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr 500 505 510	3337
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu 515 520 525	3385
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly 530 535 540	3433
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser 545 550 555 560	3481
act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile 565 570 575	3529
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt gat Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp 580 585 590	3577
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr 595 600 605	3625
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile 610 615 620	3673
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp 625 630 635 640	3721
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser 645 650 655	3769
ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala 660 665 670	3817
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly 675 680 685	3865

## 42 / 93

att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag agc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tcg gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga aag ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc	4489
Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgc gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	

## 43 / 93

atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctg	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	
gcc caa ctt gct cct ccc agc gct gct tct gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	
atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	
ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	
cag tgg atg aac cgg ctg ata gcg ttc gct tcg cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	

## 44 / 93

cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tcg tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg gta tgc acg gtg ttg act gat ttc aag acc	5353
Arg Asp Val Trp Asp Trp Val Cys Thr Val Leu Thr Asp Phe Lys Thr	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
tgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg cgg gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta gca	5881
Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala	
1345 1350 1355 1360	

## 45 / 93

gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg ccc tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	

## 46 / 93

gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tgc atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tgc gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	

## 47 / 93

cag cgg gtc gag ttc ctg gtg aat gct tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tgc ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn	
2005 2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg	
2020 2025 2030	

## 48 / 93

```

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
      2035                2040                2045

ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
      2050                2055                2060

gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
      2065                2070                2075                2080

gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
      2085                2090                2095

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
      2100                2105                2110

cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
      2115                2120                2125

gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
      2130                2135                2140

aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
      2145                2150                2155                2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
      2165                2170                2175

gcc cga ccc cgc tgg ttc acg tgg tgc cta ctc cta ctt tct gta ggg 8377
Ala Arg Pro Arg Trp Phe Thr Trp Cys Leu Leu Leu Leu Ser Val Gly
      2180                2185                2190

gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg *
      2195                2200

ggccaatagg ccattctgtt tttttccctt tttttccctt tttttttttt tttttttttt 8487
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8547
tttggtggct ccattcttagc cctagtcacg gctagctgtg aaagggtccgt gagccgcttg 8607
actgcagaga gtgctgatac tggcctctct gcagatcaag t 8648

```

&lt;210&gt; 6

&lt;211&gt; 8638

&lt;212&gt; DNA

&lt;213&gt; HCV

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1802) ... (8407)

&lt;400&gt; 6



49 / 93

```

accagccccc gattgggggc gacactccac catagatcac tccccgtgga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gactacacg gaattgccag 180
gacgaccggg tcctttcttg gatcaaccgc ctcaatgcct ggagatttgg gcgtgcccc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgoga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc agggggcgcc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggctg tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgcgc gggcaggatc tcctgtcatc tcaccttgc cctgccgaga 720
aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780
cattcgacca ccaagcgaaa catcgcatcg agcagacacg tactcggatg gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcag cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttcttg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cggttggtac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgcctttacg tatcgccgct ccgattcgc 1140
agcgcatcgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc ccagatgttaa 1200
cagaccacaa cggtttccct cttagcggat caattccgcc ccccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccgaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggctcg ttgaatgtcg tgaaggaagc 1440
agtctctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gccagaagg taccctattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gcccccgaa ccacggggac gtgggttttc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
1 5 10 15

ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
20 25 30

ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
35 40 45

caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
50 55 60

atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc 2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
65 70 75 80

aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt 2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
85 90 95

ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca 2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
100 105 110

```

## 50 / 93

tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct	2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala	
115 120 125	
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc	2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu	
130 135 140	
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg	2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val	
145 150 155 160	
gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc	2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr	
165 170 175	
tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc	2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro	
180 185 190	
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc	2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser	
195 200 205	
ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc act agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	
cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg agg cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	

## 51 / 93

ggc gac ggc agg ggg agc cta ctc tcc ccc agg ccc gtc tcc tac ttg	2857
Gly Asp Gly Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	
gac ttt gta ccc gtc gag tct atg gga acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Gly Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	
gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	
atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc	3241
Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala	
465 470 475 480	
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc	3289
Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys	
485 490 495	
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act	3337
Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr	
500 505 510	
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	

## 52 / 93

act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	
ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	
att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag agc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	

## 53 / 93

tgc gcc agg gct cag gct cca cct cca tgc tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tgc gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga aag ccg gcc atc att ccc gac agg gaa gtc ttt tac cgg gag ttc	4489
Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Phe Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgc gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	
atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	

## 54 / 93

gcc caa ctt gct cct ccc agc gct gct tct gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	
atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	
ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	
cag tgg atg aac cgg ctg ata gcg ttc gct tgc cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	
cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag acc	5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Thr	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	

## 55 / 93

cgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Arg Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg cgg gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta gca	5881
Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala	
1345 1350 1355 1360	
gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	

## 56 / 93

agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ctg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	
gag ccg ggg gat ccc gat ctc agc gac ggg cct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Pro Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac ccg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	



## 57 / 93

atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	

## 58 / 93

tgt tac ttg aag gcc gct gcg gcc	tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp		
1905	1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609	
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser		
1925	1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657	
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala		
1940	1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705	
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr		
1955	1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753	
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His		
1970	1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801	
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr		
1985	1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849	
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn		
2005	2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897	
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg		
2020	2025 2030	
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa	7945	
Met Ile Leu Met Thr His Phe Ser Ile Leu Leu Ala Gln Glu Gln		
2035	2040 2045	
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att	7993	
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile		
2050	2055 2060	
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc	8041	
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser		
2065	2070 2075 2080	
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct	8089	
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala		
2085	2090 2095	
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat	8137	
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His		
2100	2105 2110	
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct	8185	
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala		
2115	2120 2125	

## 59 / 93

```

gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc      8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
    2130                      2135                      2140

aaa ctc act cca atc ccg gct ggc tcc cag ttg gat tta tcc agc tgg      8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
    2145                      2150                      2155                      2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt      8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
                      2165                      2170                      2175

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg      8377
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Val Gly
    2180                      2185                      2190

gta ggc atc tat cta ctc ccc aac cga tga acgaggagct aaacactcca      8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg *
    2195                      2200

ggccaatagg ccatactgtt tttttttttt tttttttttt tttttttttt tttttttttt 8487
tttttttttt tttttttttt ttttttcttc ttttttttcc ttttctttcc tttggtggct 8547
ccatcttagc cctagtcacg gctagctgtg aaaggtccgt gagccgcttg actgcagaga 8607
gtgctgatac tggcctctct gcagatcaag t                                     8638

<210> 7
<211> 8638
<212> DNA
<213> HCV

<220>
<221> CDS
<222> (1802)...(8407)

<400> 7
gccagcccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggg gagtacaccg gaattgccag 180
gacgaccggg tcccttcttg gatcaaccgg ctcaatgcct ggagatttgg gcgtgcccc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaaggc ttgtggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaacc 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgcgcg cgtgttcceg ctgtcagcgc aggggcgccc gggtcttttt gtcaagaccg 540
acctgtccgg tgcctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacggcgtg tccctgcgca gctgtgctcg acgttgcac tgaagcggga agggactggc 660
tgctattggg cgaagtgcgg gggcaggatc tccctgcac tcaccttgc cctgcgagaga 720
aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780
cattcgacca ccaagcgaaa catcgcatcg agcagcacg tactcggatg gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tggtgtgggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttctcgc tgccttacgg tatcgccgct cccgattcgc 1140
agcgcatcgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200
cagaccacaa cggtttccct ctacgggat caattccgcc ccccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgcctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccgga acctggccct gtcttcttga cgagcattcc 1380

```

60 / 93

```

taggggtctt tcccctctcg ccaaaggaat gcaagggtctg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgaccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gccagaagg taccatttg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgagggttaa 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc ttgaaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
      1              5              10              15

ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
      20              25              30

ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
      35              40              45

caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
      50              55              60

atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc 2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
      65              70              75              80

aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt 2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
      85              90              95

ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca 2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
      100              105              110

tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct 2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala
      115              120              125

ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc 2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu
      130              135              140

acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg 2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val
      145              150              155              160

gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc 2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr
      165              170              175

tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc 2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro
      180              185              190

gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc 2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser
      195              200              205

```

61 / 93

ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc acc agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	
cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg aag cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Lys His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	
ggc gac agc agg ggg agc cta ctc tcc ccc cgg ccc gtc tcc tac ttg	2857
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	
gac ttt gta ccc gtc gag tct atg gaa acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	

## 62 / 93

gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	
atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc	3241
Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala	
465 470 475 480	
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc	3289
Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys	
485 490 495	
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act	3337
Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr	
500 505 510	
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	
act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgc gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	

## 63 / 93

ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tgc cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	
att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag ggc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tcg gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gct gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	

## 64 / 93

gga agg ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc	4489
Gly Arg Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgt gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	
atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Pro Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	
gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	
atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	
ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	



## 65 / 93

cag tgg atg aac cgg ctg ata gcg ttc gct tgc cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	
cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag gcc	5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Ala	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
tgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg cga gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg ccg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	

## 66 / 93

aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ccg gtt ggg tca cag ctc cca tgc gag ccc gaa ctg gac gta gca	5881
Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Leu Asp Val Ala	
1345 1350 1355 1360	
gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gag aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg ccg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	

## 67 / 93

agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	
gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	

## 68 / 93

gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	

## 69 / 93

```

ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat 7849
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn
                2005                2010                2015

tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg 7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg
                2020                2025                2030

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
                2035                2040                2045

ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
                2050                2055                2060

gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
2065                2070                2075                2080

gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
                2085                2090                2095

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
                2100                2105                2110

cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
                2115                2120                2125

gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
                2130                2135                2140

aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
2145                2150                2155                2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
                2165                2170                2175

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 8377
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly
                2180                2185                2190

gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg *
                2195                2200

ggccaatagg ccacccctgtt ttttccctt tttttttttt tttttttttt tttttttttt 8487
tttttttttt ttttccctt ttttttttct cttttttttcc ttttctttcc tttggtggct 8547
ccatcttagc cctagtcacg gctagctgtg aaaggctcgt gagccgcttg actgcagaga 8607
gtgctgatac tggcctctct gcagatcaag t 8638

```

70 / 93

<211> 6  
<212> DNA  
<213> HCV

<400> 8  
accagc

6

<210> 9  
<211> 63  
<212> DNA  
<213> HCV

<400> 9  
gaattccaga tggcgcgccc agatgttaac cagatccatg gcacactcta gagtactgtc 60  
gac 63

<210> 10  
<211> 33  
<212> DNA  
<213> HCV

<400> 10  
cggaatcggtt aacagaccac aacggttttcc ctc 33

<210> 11  
<211> 30  
<212> DNA  
<213> HCV

<400> 11  
ggcgtaacca tgggtattatc gtgtttttcca 30

<210> 12  
<211> 45  
<212> DNA  
<213> HCV

<400> 12  
gcatatgaat tctaatacga ctactatag gccagcccc gattg 45

<210> 13  
<211> 45  
<212> DNA  
<213> HCV

<400> 13  
ggcgcgccct ttggtttttc tttgaggttt aggattcgtg ctcac 45

<210> 14  
<211> 36  
<212> DNA  
<213> HCV

<400> 14  
aaagggcgca tgattgaaca agatggattg cacgca 36

<210> 15  
<211> 39

71 / 93

<212> DNA  
<213> HCV

<400> 15  
gcatatgtta actcagaaga actcgtcaag aaggcgata 39

<210> 16  
<211> 45  
<212> DNA  
<213> HCV

<400> 16  
gcatatgaat tctaatacga ctactatag gccagcccc gattg 45

<210> 17  
<211> 30  
<212> DNA  
<213> HCV

<400> 17  
acgcagaaag cgtctagcca tggcgtagt 30

<210> 18  
<211> 30  
<212> DNA  
<213> HCV

<400> 18  
tcccggggca ctcgcaagca ccctatcagg 30

<210> 19  
<211> 26  
<212> DNA  
<213> HCV

<220>  
<223> Label with FAM: fluorescence reporter dye  
<223> Label with TAMRA: Quencher dye

<400> 19  
tggtctgcgg aacgggtgag tacacc 26

<210> 20  
<211> 45  
<212> DNA  
<213> HCV

<400> 20  
gtggacgaat tctaatacga ctactataa ccagccccg attgg 45

<210> 21  
<211> 27  
<212> DNA  
<213> HCV

<400> 21  
ggaacgcccg tcgtggccag ccacgat 27

72 / 93

<210> 22  
 <211> 23  
 <212> DNA  
 <213> HCV

<400> 22  
 gtcgtcttct ctgacatgga gac

23

<210> 23  
 <211> 27  
 <212> DNA  
 <213> HCV

<400> 23  
 gagttgctca gtggattgat gggcagc

27

<210> 24  
 <211> 8638  
 <212> DNA  
 <213> HCV

<220>  
 <221> CDS  
 <222> (1802) ... (8407)

<400> 24  
 accagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60  
 tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgctcgtcag cctccaggac 120  
 cccccctccc gggagagcca tagtggtctg cggaaccggg gagtacaccg gaattgccag 180  
 gacgaccggg tcctttcttg gatcaaccgc ctcaatgcct ggagatttgg gcgtgcccc 240  
 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gctgatagg 300  
 gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360  
 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420  
 cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480  
 ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc gggtcttttt gtcaagaccg 540  
 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600  
 cgacgggcgt tccttgccga gctgtgctcg acgttgtcac tgaagcggga agggactggc 660  
 tgctattggg cgaagtgcg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720  
 aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780  
 cattcgacca ccaagcgaac catcgcatcg agcagcacg tactcggtat gaagccggtc 840  
 ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900  
 ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960  
 gcttgccgaa tatcatggtg gaaaatggcc gcttttcttg attcatcgac tgtggccggc 1020  
 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080  
 ttggcggcga atgggctgac cgcttctctg tgctttacgg tatcgccgct cccgattcgc 1140  
 agcgcacatgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200  
 cagaccacaa cggtttccct cttagcggggt caattccgcc ccccccccta acgttactgg 1260  
 ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320  
 gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380  
 taggggtctt tcccctctcg ccaaagggaat gcaaggctctg ttgaatgtcg tgaaggaaac 1440  
 agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500  
 gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560  
 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620  
 atggctctcc tcaagcgtat tcaacaagg gctgaaggat gccagaagg taccctattg 1680  
 tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740  
 aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800  
 c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849



73 / 93

Met	Asp	Arg	Glu	Met	Ala	Ala	Ser	Cys	Gly	Gly	Ala	Val	Phe	Val	Gly	
1				5					10						15	
ctg	ata	ctc	ttg	acc	ttg	tca	cgc	cac	tat	aag	ctg	ttc	ctc	gct	agg	1897
Leu	Ile	Leu	Leu	Thr	Leu	Ser	Pro	His	Tyr	Lys	Leu	Phe	Leu	Ala	Arg	
			20					25					30			
ctc	ata	tgg	tgg	tta	caa	tat	ttt	atc	acc	agg	gcc	gag	gca	cac	ttg	1945
Leu	Ile	Trp	Trp	Leu	Gln	Tyr	Phe	Ile	Thr	Arg	Ala	Glu	Ala	His	Leu	
		35				40					45					
caa	gtg	tgg	atc	ccc	ccc	ctc	aac	gtt	cgg	ggg	ggc	cgc	gat	gcc	gtc	1993
Gln	Val	Trp	Ile	Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Val	
	50					55					60					
atc	ctc	ctc	acg	tgc	gcg	atc	cac	cca	gag	cta	atc	ttt	acc	atc	acc	2041
Ile	Leu	Leu	Thr	Cys	Ala	Ile	His	Pro	Glu	Leu	Ile	Phe	Thr	Ile	Thr	
65					70				75					80		
aaa	atc	ttg	ctc	gcc	ata	ctc	ggc	cca	ctc	atg	gtg	ctc	cag	gct	ggc	2089
Lys	Ile	Leu	Leu	Ala	Ile	Leu	Gly	Pro	Leu	Met	Val	Leu	Gln	Ala	Gly	
				85				90						95		
ata	acc	aaa	gtg	cgc	tac	ttc	gtg	cgc	gca	cac	ggg	ctc	att	cgt	gca	2137
Ile	Thr	Lys	Val	Pro	Tyr	Phe	Val	Arg	Ala	His	Gly	Leu	Ile	Arg	Ala	
			100					105					110			
tgc	atg	ctg	gtg	cgc	aag	gtt	gct	ggg	ggc	cat	tat	gtc	caa	atg	gct	2185
Cys	Met	Leu	Val	Arg	Lys	Val	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	
		115					120					125				
ctc	atg	aag	ttg	gcc	gca	ctg	aca	ggc	acg	tac	gtt	tat	gac	cat	ctc	2233
Leu	Met	Lys	Leu	Ala	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Asp	His	Leu	
	130					135					140					
acc	cca	ctg	cgc	gac	tgg	gcc	cac	gcg	ggc	cta	cga	gac	ctt	gcg	gtg	2281
Thr	Pro	Leu	Arg	Asp	Trp	Ala	His	Ala	Gly	Leu	Arg	Asp	Leu	Ala	Val	
145					150				155					160		
gca	gtt	gag	ccc	gtc	gtc	ttc	tct	gat	atg	gag	acc	aag	gtt	atc	acc	2329
Ala	Val	Glu	Pro	Val	Val	Phe	Ser	Asp	Met	Glu	Thr	Lys	Val	Ile	Thr	
			165					170					175			
tgg	ggg	gca	gac	acc	gcg	gcg	tgt	ggg	gac	atc	atc	ttg	ggc	ctg	ccc	2377
Trp	Gly	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Leu	Gly	Leu	Pro	
		180					185						190			
gtc	tcc	gcc	cgc	agg	ggg	agg	gag	ata	cat	ctg	gga	ccg	gca	gac	agc	2425
Val	Ser	Ala	Arg	Arg	Gly	Arg	Glu	Ile	His	Leu	Gly	Pro	Ala	Asp	Ser	
		195				200					205					
ctt	gaa	ggg	cag	ggg	tgg	cga	ctc	ctc	gcg	cct	att	acg	gcc	tac	tcc	2473
Leu	Glu	Gly	Gln	Gly	Trp	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ser	
	210					215					220					
caa	cag	acg	cga	ggc	cta	ctt	ggc	tgc	atc	atc	act	agc	ctc	aca	ggc	2521
Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly	
225					230				235						240	

74 / 93

cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg agg cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	
ggc gac agc agg ggg agc cta ctc tcc ccc agg ccc gtc tcc tac ttg	2857
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	
gac ttt gta ccc gtc gag tct atg gaa acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	
gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	

## 75 / 93

atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc	3241
Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala	
465 470 475 480	
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc	3289
Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys	
485 490 495	
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act	3337
Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr	
500 505 510	
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	
act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgt gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	
ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	

## 76 / 93

att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag agc gtc	4057
Pro Gly Leu Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	
tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tcg gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gcc gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga aag ccg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc	4489
Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgc gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	

## 77 / 93

atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctg	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	
gcc caa ctt gct cct ccc agc gct gct tct gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	
atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	
ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	
cag tgg atg aac cgg ctg ata gcg ttc gct tcg cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	

## 78 / 93

cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tcg tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag acc	5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Thr	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg cgg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
ggg tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Gly Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg cgg gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta gca	5881
Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala	
1345 1350 1355 1360	

## 79 / 93

gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc ccg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gaa aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	
cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	

## 80 / 93

gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tgc atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	



## 81 / 93

cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	
tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn	
2005 2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg	
2020 2025 2030	

## 82 / 93

```

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
      2035                      2040                      2045

ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
      2050                      2055                      2060

gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
      2065                      2070                      2075                      2080

gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
      2085                      2090                      2095

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
      2100                      2105                      2110

cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
      2115                      2120                      2125

gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
      2130                      2135                      2140

aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
      2145                      2150                      2155                      2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
      2165                      2170                      2175

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 8377
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly
      2180                      2185                      2190

gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg *
      2195                      2200

ggccaatagg ccatcctggt tttttccctt tttttttttt tttttttttt tttttttttt 8487
tttttttttt tttttctcct tttttttcct cttttttttcc tttttctttcc tttgggtggct 8547
ccatccttagc cctagtcacg gctagctgtg aaagggtccgt gagccgcttg actgcagaga 8607
gtgctgatac tggcctctct gcagatcaag t 8638

```

&lt;210&gt; 25

&lt;211&gt; 8638

&lt;212&gt; DNA

&lt;213&gt; HCV

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1802)...(8407)

83 / 93

&lt;400&gt; 25

```

accagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggctcg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg tcctttcttg gatcaaccgg ctcaatgcct ggagatttgg gcgtgcccc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttcagg ctgtcagcgc agggcgcccc gggtcttttt gtcaagaccg 540
acctgtccgg tgcctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacggcggt tccttgccga gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgcgc gggcaggatc tcctgtcatc tcacctgtct cctgccgaga 720
aagtatccat catggtgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780
cattcgacca ccaagcgaac catcgcatcg agcgagcacg tactcggtat gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggccgcgatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttcttg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcgcgga atgggctgac cgcttcctcg tgccttacgg tatcgccgct cccgattcgc 1140
agcgcacgac cttctatcgc cttcttgacg agttctcttg agttcgcgcc cagatgttaa 1200
cagaccacaa cggtttcctc ctacggggat caattccgcc ccccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccgaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggctcg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaaggc gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaagg gctgaaggat gccagaagg taccatttg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
1 5 10 15

```

```

ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
20 25 30

```

```

ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
35 40 45

```

```

caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
50 55 60

```

```

atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc 2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
65 70 75 80

```

```

aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt 2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
85 90 95

```

```

ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca 2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
100 105 110

```

84 / 93

tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct	2185
Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala	
115 120 125	
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc	2233
Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu	
130 135 140	
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg	2281
Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val	
145 150 155 160	
gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc	2329
Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr	
165 170 175	
tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc	2377
Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro	
180 185 190	
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc	2425
Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser	
195 200 205	
ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac tcc	2473
Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser	
210 215 220	
caa cag acg cga ggc cta ctt ggc tgc atc atc acc agc ctc aca ggc	2521
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
225 230 235 240	
cgg gac agg aac cag gtc gag ggg gag gtc caa gtg gtc tcc acc gca	2569
Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala	
245 250 255	
aca caa tct ttc ctg gcg acc tgc gtc aat ggc gtg tgt tgg act gtc	2617
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val	
260 265 270	
tat cat ggt gcc ggc tca aag acc ctt gcc ggc cca aag ggc cca atc	2665
Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile	
275 280 285	
acc caa atg tac acc aat gtg gac cag gac ctc gtc ggc tgg caa gcg	2713
Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala	
290 295 300	
ccc ccc ggg gcg cgt tcc ttg aca cca tgc acc tgc ggc agc tcg gac	2761
Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
305 310 315 320	
ctt tac ttg gtc acg aag cat gcc gat gtc att ccg gtg cgc cgg cgg	2809
Leu Tyr Leu Val Thr Lys His Ala Asp Val Ile Pro Val Arg Arg Arg	
325 330 335	

## 85 / 93

ggc gac agc agg ggg agc cta ctc tcc ccc cgg ccc gtc tcc tac ttg	2857
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu	
340 345 350	
aag ggc tct tcg ggc ggt cca ctg ctc tgc ccc tcg ggg cac gct gtg	2905
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val	
355 360 365	
ggc atc ttt cgg gct gcc gtg tgc acc cga ggg gtt gcg aag gcg gtg	2953
Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
370 375 380	
gac ttt gta ccc gtc gag tct atg gaa acc act atg cgg tcc ccg gtc	3001
Asp Phe Val Pro Val Glu Ser Met Glu Thr Met Arg Ser Pro Val	
385 390 395 400	
ttc acg gac aac tcg tcc cct ccg gcc gta ccg cag aca ttc cag gtg	3049
Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val	
405 410 415	
gcc cat cta cac gcc cct act ggt agc ggc aag agc act aag gtg ccg	3097
Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	
420 425 430	
gct gcg tat gca gcc caa ggg tat aag gtg ctt gtc ctg aac ccg tcc	3145
Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser	
435 440 445	
gtc gcc gcc acc cta ggt ttc ggg gcg tat atg tct aag gca cat ggt	3193
Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly	
450 455 460	
atc gac cct aac atc aga acc ggg gta agg acc atc acc acg ggt gcc	3241
Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala	
465 470 475 480	
ccc atc acg tac tcc acc tat ggc aag ttt ctt gcc gac ggt ggt tgc	3289
Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys	
485 490 495	
tct ggg ggc gcc tat gac atc ata ata tgt gat gag tgc cac tca act	3337
Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr	
500 505 510	
gac tcg acc act atc ctg ggc atc ggc aca gtc ctg gac caa gcg gag	3385
Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu	
515 520 525	
acg gct gga gcg cga ctc gtc gtg ctc gcc acc gct acg cct ccg gga	3433
Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly	
530 535 540	
tcg gtc acc gtg cca cat cca aac atc gag gag gtg gct ctg tcc agc	3481
Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser	
545 550 555 560	

## 86 / 93

act gga gaa atc ccc ttt tat ggc aaa gcc atc ccc atc gag acc atc	3529
Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile	
565 570 575	
aag ggg ggg agg cac ctc att ttc tgc cat tcc aag aag aaa tgc gat	3577
Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp	
580 585 590	
gag ctc gcc gcg aag ctg tcc ggc ctc gga ctc aat gct gta gca tat	3625
Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr	
595 600 605	
tac cgg ggc ctt gat gta tcc gtc ata cca act agc gga gac gtc att	3673
Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile	
610 615 620	
gtc gta gca acg gac gct cta atg acg ggc ttt acc ggc gat ttc gac	3721
Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp	
625 630 635 640	
tca gtg atc gac tgc aat aca tgt gtc acc cag aca gtc gac ttc agc	3769
Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser	
645 650 655	
ctg gac ccg acc ttc acc att gag acg acg acc gtg cca caa gac gcg	3817
Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala	
660 665 670	
gtg tca cgc tcg cag cgg cga ggc agg act ggt agg ggc agg atg ggc	3865
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly	
675 680 685	
att tac agg ttt gtg act cca gga gaa cgg ccc tcg ggc atg ttc gat	3913
Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp	
690 695 700	
tcc tcg gtt ctg tgc gag tgc tat gac gcg ggc tgt gct tgg tac gag	3961
Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu	
705 710 715 720	
ctc acg ccc gcc gag acc tca gtt agg ttg cgg gct tac cta aac aca	4009
Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr	
725 730 735	
cca ggg ttg ccc gtc tgc cag gac cat ctg gag ttc tgg gag ggc gtc	4057
Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly Val	
740 745 750	
ttt aca ggc ctc acc cac ata gac gcc cat ttc ttg tcc cag act aag	4105
Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys	
755 760 765	
cag gca gga gac aac ttc ccc tac ctg gta gca tac cag gct acg gtg	4153
Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val	
770 775 780	
tgc gcc agg gct cag gct cca cct cca tcg tgg gac caa atg tgg aag	4201
Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys	
785 790 795 800	

## 87 / 93

tgt ctc ata cgg cta aag cct acg ctg cac ggg cca acg ccc ctg ctg	4249
Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu	
805 810 815	
tat agg ctg gga gcc gtt caa aac gag gtt act acc aca cac ccc ata	4297
Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr Thr His Pro Ile	
820 825 830	
acc aaa tac atc atg gca tgc atg tcg gct gac ctg gag gtc gtc acg	4345
Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr	
835 840 845	
agc acc tgg gtg ctg gta ggc gga gtc cta gca gct ctg gct gcg tat	4393
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr	
850 855 860	
tgc ctg aca aca ggc agc gtg gtc att gtg ggc agg atc atc ttg tcc	4441
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser	
865 870 875 880	
gga agg ccg gcc atc att ccc gac agg gaa gtc ctt tac ccg gag ttc	4489
Gly Arg Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe	
885 890 895	
gat gag atg gaa gag tgt gcc tca cac ctc cct tac atc gaa cag gga	4537
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly	
900 905 910	
atg cag ctc gcc gaa caa ttc aaa cag aag gca atc ggg ttg ctg caa	4585
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln	
915 920 925	
aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag	4633
Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys	
930 935 940	
tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc	4681
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile	
945 950 955 960	
agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc	4729
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro	
965 970 975	
gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc	4777
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu	
980 985 990	
acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc	4825
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala	
995 1000 1005	
gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc	4873
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly	
1010 1015 1020	

## 88 / 93

atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg	4921
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val	
1025 1030 1035 1040	
gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc	4969
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala	
1045 1050 1055	
ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac	5017
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn	
1060 1065 1070	
cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg	5065
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val	
1075 1080 1085	
tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg	5113
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	
1090 1095 1100	
cag tgg atg aac cgg ctg ata gcg ttc gct tgc cgg ggt aac cac gtc	5161
Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val	
1105 1110 1115 1120	
tcc ccc acg cac tat gtg cct gag agc gac gct gca gca cgt gtc act	5209
Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr	
1125 1130 1135	
cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac	5257
Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His	
1140 1145 1150	
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tgc tgg cta	5305
Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu	
1155 1160 1165	
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag gcc	5353
Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Ala	
1170 1175 1180	
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc	5401
Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe	
1185 1190 1195 1200	
tca tgt caa cgt ggg tac aag gga gtc tgg ccg ggc gac ggc atc atg	5449
Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met	
1205 1210 1215	
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac	5497
Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn	
1220 1225 1230	
tgt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat	5545
Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His	
1235 1240 1245	



## 89 / 93

gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc	5593
Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser	
1250 1255 1260	
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gag	5641
Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu	
1265 1270 1275 1280	
tac gtg gag gtt acg cga gtg ggg gat ttc cac tac gtg acg ggc atg	5689
Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met	
1285 1290 1295	
acc act gac aac gta aag tgc cgg tgt cag gtt ccg gcc ccc gaa ttc	5737
Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe	
1300 1305 1310	
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc	5785
Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys	
1315 1320 1325	
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa	5833
Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln	
1330 1335 1340	
tac ccg gtt ggg tca cag ctc cca tgc gag ccc gaa ctg gac gta gca	5881
Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Leu Asp Val Ala	
1345 1350 1355 1360	
gtg ctc act tcc atg ctc acc gac ccc tcc cac att acg gcg gag acg	5929
Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr	
1365 1370 1375	
gct aag cgt agg ctg gcc agg gga tct ccc ccc tcc ttg gcc agc tca	5977
Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser	
1380 1385 1390	
tca gct agc cag ctg tct gcg cct tcc ttg aag gca aca tgc act acc	6025
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr	
1395 1400 1405	
cgt cat gac tcc cgg gac gct gac ctc atc gag gcc aac ctc ctg tgg	6073
Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp	
1410 1415 1420	
cgg cag gag atg ggc ggg aac atc acc cgc gtg gag tca gag aat aag	6121
Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys	
1425 1430 1435 1440	
gta gta att ttg gac tct ttc gag ccg ctc caa gcg gag gag gat gag	6169
Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu	
1445 1450 1455	
agg gaa gta tcc gtt ccg gcg gag atc ctg cgg agg tcc agg aaa ttc	6217
Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe	
1460 1465 1470	

## 90 / 93

cct cga gcg atg ccc ata tgg gca cgc ccg gat tac aac cct cca ctg	6265
Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
1475 1480 1485	
tta gag tcc tgg aag gac ccg gac tac gtc cct cca gtg gta cac ggg	6313
Leu Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly	
1490 1495 1500	
tgt cca ttg ccg cct gcc aag gcc cct ccg ata cca cct cca cgg agg	6361
Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg	
1505 1510 1515 1520	
aag agg acg gtt gtc ctg tca gaa tct acc gtg tct tct gcc ttg gcg	6409
Lys Arg Thr Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala	
1525 1530 1535	
gag ctc gcc aca aag acc ttc ggc agc tcc gaa tcg tcg gcc gtc gac	6457
Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp	
1540 1545 1550	
agc ggc acg gca acg gcc tct cct gac cag ccc tcc gac gac ggc gac	6505
Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp	
1555 1560 1565	
gcg gga tcc gac gtt gag tcg tac tcc tcc atg ccc ccc ctt gag ggg	6553
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
1570 1575 1580	
gag ccg ggg gat ccc gat ctc agc gac ggg tct tgg tct acc gta agc	6601
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
1585 1590 1595 1600	
gag gag gct agt gag gac gtc gtc tgc tgc tcg atg tcc tac aca tgg	6649
Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
1605 1610 1615	
aca ggc gcc ctg atc acg cca tgc gct gcg gag gaa acc aag ctg ccc	6697
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro	
1620 1625 1630	
atc aat gca ctg agc aac tct ttg ctc cgt cac cac aac ttg gtc tat	6745
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr	
1635 1640 1645	
gct aca aca tct cgc agc gca agc ctg ccg cag aag aag gtc acc ttt	6793
Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe	
1650 1655 1660	
gac aga ctg cag gtc ctg gac gac cac tac ccg gac gtg ctc aag gag	6841
Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu	
1665 1670 1675 1680	
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag	6889
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu	
1685 1690 1695	

## 91 / 93

gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc	6937
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly	
1700 1705 1710	
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac	6985
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His	
1715 1720 1725	
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att	7033
Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile	
1730 1735 1740	
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag	7081
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu	
1745 1750 1755 1760	
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg	7129
Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly	
1765 1770 1775	
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc	7177
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu	
1780 1785 1790	
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga	7225
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly	
1795 1800 1805	
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct	7273
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro	
1810 1815 1820	
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag	7321
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu	
1825 1830 1835 1840	
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc	7369
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala	
1845 1850 1855	
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc	7417
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile	
1860 1865 1870	
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg	7465
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg	
1875 1880 1885	
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr	
1890 1895 1900	
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac	7561
Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	
1905 1910 1915 1920	

## 92 / 93

tgc acg atg ctc gta tgc gga gac gac ctt gtc gtt atc tgt gaa agc	7609
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser	
1925 1930 1935	
gcg ggg acc caa gag gac gag gcg agc cta cgg gcc ttc acg gag gct	7657
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala	
1940 1945 1950	
atg act aga tac tct gcc ccc cct ggg gac ccg ccc aaa cca gaa tac	7705
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr	
1955 1960 1965	
gac ttg gag ttg ata aca tca tgc tcc tcc aat gtg tca gtc gcg cac	7753
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His	
1970 1975 1980	
gat gca tct ggc aaa agg gtg tac tat ctc acc cgt gac ccc acc acc	7801
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr	
1985 1990 1995 2000	
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtc aat	7849
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn	
2005 2010 2015	
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca agg	7897
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg	
2020 2025 2030	
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa	7945
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln	
2035 2040 2045	
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att	7993
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile	
2050 2055 2060	
gag cca ctt gac cta cct cag atc att caa cga ctc cac gcc ctt agc	8041
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser	
2065 2070 2075 2080	
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct	8089
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala	
2085 2090 2095	
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat	8137
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His	
2100 2105 2110	
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct	8185
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala	
2115 2120 2125	
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc	8233
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu	
2130 2135 2140	

## 93 / 93

```

aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg      8281
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
2145                2150                2155                2160

ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt      8329
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
                2165                2170                2175

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg      8377
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly
                2180                2185                2190

gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactóca      8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg  *
                2195                2200

ggccaatagg ccatactggt tttttccctt tttttttttt tttttttttt tttttttttt 8487
tttttttttt ttttctcctt tttttttcct cttttttttcc ttttctttcc tttggtggct 8547
ccatcttagc cctagtcacg gctagctgtg aaaggtcogt gagccgcttg actgcagaga 8607
gtgctgatac tggcctctct gcagatcaag t                                8638

```